

Федеральное государственное бюджетное образовательное учреждение
высшего образования «Санкт-Петербургский государственный университет»
(СПбГУ)

**АКТУАЛЬНЫЕ ПРОБЛЕМЫ ФОРМИРОВАНИЯ
«МАШИННОЙ ЭТИКИ»**

Выпускная квалификационная работа по направлению

47.03.02. «Прикладная этика»

основная образовательная программа СВ.5043.2014 «Прикладная этика»

Исполнитель

Кутузова Анна Олеговна

Научный руководитель

к.филос.н., доцент

Перов Вадим Юрьевич

Рецензент

д.филос.н., доцент

Рукавишников Алексей Борисович

Санкт-Петербург

2018

СОДЕРЖАНИЕ

Введение.....	3-8
Глава 1 Машинная этика.....	9-20
1.1 Определение машинной этики.....	9
1.2 Значение принципов робототехники для машинной этики.....	10
1.3 Определение робота.....	13
1.4 Сильный и слабый искусственный интеллект.....	17
Глава 2 Теоретические проблемы машинной этики.....	21-34
2.1 Искусственный моральный агент.....	21
2.2 Мораль искусственного интеллекта.....	27
Глава 3 Прикладные проблемы машинной этики.....	35-61
3.1 Понятие и виды боевого робота.....	35
3.2 Боевые роботы и проблемы машинной этики.....	37
3.3 Понятие социального робота.....	42
3.4 Роботы в медицине.....	43
3.5 Здравоохранение и роботы ассистенты.....	46
Заключение.....	62
Список литературы.....	63-65

Введение

Искусственный интеллект – это сложная тема, пользующаяся большим интересом среди учёных технических и гуманитарных областей. Она очень любопытна в связи с несколькими особенностями.

Первой особенностью является относительно недавнее открытие такого масштабного поля для научных исследований. Первые разработки роботов, которые стали прообразами современных моделей, были начаты в 30-е годы XX века. О нейронных сетях научное сообщество начинало узнавать в 40-е года XX века. И уже к концу XX века и по сей день человек имеет возможность взаимодействовать с роботами, например с роботом-пылесосом в своей квартире.

Второй особенностью привлекательности этой темы является её поступательное развитие. Открытия новых решений даёт толчок в развитии индустрии робототехники, т.е. увеличивает сложность изготавливаемых систем и машин, а так же количество создаваемых продуктов и сферу распространения. Таким образом, машина с ИИ становится частью социальной реальности людей. Группы людей, отдельные лица выстраивают взаимодействие с машиной.

Третья особенность, которая очень важна для этического ракурса изучения проблемы - это поворот от понимания робота, как развлекательной продукции, к направлению разработок и исследований социального взаимодействия робота с человеком.

В связи с этим аспектом возрастает необходимость в развитии этики искусственного интеллекта, которая занимается обнаружением проблем и поиском решений. Однако в этике ИИ произошло деление на две сферы - робоэтику и машинную этику. Работа посвящена проблемам формирования машинной этики. Ведь она задаётся вопросами: каким образом машина может влиять на человека, каким образом можно реализовать механизм принятия этических решений машиной? Как машина меняет границы

этически дозволенного и понимание ценностных категорий? Каким образом робот может воспринимать нас при социальном взаимодействии и как он может быть нравственным? Это одна из сторон взаимодействия машины и человека, которая менее исследована, а поиск и нахождение решений интересен инженерам и разработчикам, профессионалам, которые будут взаимодействовать с ними и обычным пользователям робототехники.

Одними из авторитетных авторов, занимающихся машинной этикой, повлиявших на написание этой работы являются Джеймс Мур, который предложил научному сообществу типологию моральных агентов и развивает теоретическое направление машинной этики. Так же среди учёных, оказавших влияние нужно выделить группу из К.Аллена, В.Валлаха и И.Смит, потому что их исследования в области искусственной морали являются актуальными и позволяют смотреть на современное состояние взаимодействия этики и технологий.

Среди основных источников необходимо отметить ежегодный отчёт по робоэтике сделанный в этическом комитете ЮНЕСКО в 2016 и 2017 годах, т.к. сравнение отчётов дало возможность проанализировать развитие и модификации этических и технологических решений и выявить прикладные проблемы в специализированных областях медицины и обороны.

Помимо официальных документов, законов и ГОСТов для формирования выводов были взяты во внимание художественная литература и кинематограф, которые повлияли на формирование понятий «робот» и других, которыми пользуется научное сообщество. Они отобразили свой собственный взгляд на то, как выглядит нравственный робот в произведениях Айзека Азимова и заставили задуматься над угрозами глобальных проблем, например в фильме «Терминатор 3: Восстание машин».

Целью работы является представить текущий этап развития одного из направлений этики ИИ, представить современное понимание проблем теоретического и прикладного характера, которые проявляются в обществе с приходом новых технологий и формируют облик машинной этики.

Задачи, поставленные в работе заключались в том, чтобы

- найти не устаревшие научные источники для формирования выводов о сфере, которая постоянно находится в развитии
- дать определение главным понятиям работы – робот, ИИ, машинная этика
- показать место машинной этики среди других направлений, показать и теоретические аспекты, и прикладные
- проанализировать найденные проблемы, попытаться дать этические рекомендации для их решения.

В итоге для достижения цели исследования были сделаны такие шаги:

- Сформулировано понятие машинной этики, выявлено её положение и отношение к робоэтике.
- Даны определения: что такое ИИ, робот, социальный робот и военный робот.
- Найдены проблемы, которые затрагивают теоретический слой, проанализированы с помощью применения на них этических теорий.
- Выяснено, какие проблемы возникают у людей, которые непосредственно имеют дело с социальной робототехникой и военной.

По окончании исследования были сделаны следующие выводы.

Машинная этика – это область моральной философии, которая занимается разработкой подходов к установлению механизма принятия этических решений интеллектуальными машинами и системами. Это необходимо для взаимодействия человека и искусственного морального агента. Машинная этика имеет междисциплинарный характер и ориентируется на разработки робототехники и других компьютерных наук. В русской литературе, и научной, и публицистической в большинстве случаев происходит подмена машинной этики робоэтикой или их смешивание, таким образом, в обыденном сознании закрепляется смешанный образ этих областей. Эти области представляют направление этики ИИ, но робоэтика направлена на

регуляцию поведения и отношений со стороны людей к машинам, а машинная этика занята вопросом о том, каким образом машина будет относиться к людям.

Выяснилось, что под роботом следует понимать механизм, с определённым набором компонентов – это электромеханическое устройство, способность вступать во взаимодействие с окружающей средой, мобильность устройства, степень свободы, программируемость, автономность. Кроме роботов было необходимо выявить понятие искусственного интеллекта, удалось установить то, что ИИ – это одновременно и научная сфера, и способность. ИИ бывает слабым не имеет полной автономии, но ведутся дискуссии о возможности ИИ тождественного человеческому, сильного. Это позволяет сделать вывод о том, что машинная этика имеет дело только со слабым ИИ, который не может быть полностью автономным.

Выводы о моральном агенте можно представить так: сейчас работа машинной этики возможно только с тремя из 4 возможным видов моральных агентов, а моральный агент – это действующее лицо, способное выражать моральную оценку, опирающееся на моральные нормы, образцы и этические категории в своей деятельности. Два подхода к искусственной морали, директивный, внедрение правил, и восходящий, обучающий, не подходят для воплощения. При применении на них различных этических теорий возникают затруднения воплощения и в первом, и втором подходе: слабая степень апробации подходов в реальности, невозможность единого метода обучения роботов, недостаточное развитие робототехники на данный момент.

Последняя часть даёт право сделать такие выводы: актуальные проблемы прикладного характера на примере социальной и боевой робототехники сходятся в вопросах безопасности, конфиденциальности и защиты данных, этические проблемы обеих сфер касались ценности человеческой жизни, уважения и достоинства пользователя робототехники. Проблемы боевых роботов связаны с акцентом на насильственную деятельность и её деструктивный характер, проблемы компьютеризации

войны и легитимации прав на убийство это показывают. Социальная робототехника акцентирует внимание на лицах, попадающих в уязвимые группы: детей, пенсионеров. В проблемах изоляции и самообмана, связанных с роботами в медицине, учёные переходят в контекст, который больше связан с психологией, а не с машинной этикой. Шаги, которые помогут встать на путь решения проблем обеих сфер заключаются в том, чтобы наладить совместную работу учёных инженеров и этиков, чтобы устранить все возможные риски и угрозы. Второй шаг – это формулирование принципов робототехники и модификация международных документов, например Женевской конференции о защите прав на войне, и информированного согласия пациентов.

Первая глава исследования посвящена определению машинной этики. Было проанализировано, что машинная этика вступает во взаимосвязь с робоэтикой в параграфе 1.2. Так же параграфы 1.3 и 1.4 обращён к понятию робота и пониманию искусственного интеллекта. Были выявлены характеристики робота и два понимания ИИ, как науки, и как способности.

Вторая глава посвящена теоретическим проблемам, с которыми столкнулось научное сообщество. В параграфе 2.1. рассмотрены проблемы определения морального агента, представлены типы агентов и проанализированы их возможности. Вторая проблема рассмотренная в параграфе 2.2. касалась механизма искусственной морали и выявлены два подхода – директивный и восходящий. Слабые и сильные стороны были проанализированы с помощью применения концепции утилитаризма и деонтологии. В результате было установлено определение морального агента и причины из-за которых нельзя отдать предпочтение одному из подходов искусственной морали.

Третья глава работы фокусируется на нескольких сферах социальной и военной, которые не только по прогнозам футурологов, но и по активному внедрению компьютеров с ИИ и роботов могут отразить насущные проблемы

машинной этики в реальных условиях использования этих устройств. В ходе работы были затронуты проблемы безопасности, конфиденциальности для тех и других сфер. Военная робототехника представила проблемы компьютеризации войны, легитимации права робота на убийство, проблему достоинства людей, использующих роботов для ведения военных действий, они отображены в параграфах 3.1 и 3.2.

Социальная робототехника была рассмотрена на примере робота, участвующего в терапии пожилых людей со слабоумием, детей с аутизмом как представителей одних из наиболее уязвимых групп населения. В результате разворачивания мысли в параграфах 3.3-3.5 выявились проблемы самообмана о роботе, как живом существе среди детей и пенсионеров, проблемы изоляции больных при проведении терапии, проблемы защиты данных и достоинства больных, которых нужно лечить.

Апробация работы проводилась на начальных этапах её формирования. Для того чтобы ознакомиться с современным состоянием и взглядами молодых учёных на поднимаемую проблему, был сделан доклад «Актуальные проблемы формирования машинной этики» и продемонстрирован в Санкт-Петербургском государственном университете на девятой международной этической конференции: «Теоретическая и прикладная этика: традиции и перспективы – 2017». Он был представлен на студенческой секции, которая объединила молодых учёных, темой революционности технологий. Выступление показало, какие вопросы работы требуется исследовать дальше, в каком направлении выстраивать дальнейшую работу.

Работа состоит из 3 глав, 11 параграфов, введения, заключения и списка литературы из 29 наименований, большинство из которых являются иностранными источниками.

Глава 1 Машинная этика

Определение машинной этики

В большом разделе этики, посвященному появлению, развитию и проблемам искусственного интеллекта (ИИ), за всё время научных изысканий успело выделиться и сформироваться два подраздела: робоэтика и машинная этика.

Машинная этика – область исследований моральной философии, которая занимается поиском ответов на вопрос: как реализовать механизм принятия этических решений в роботах, системах или компьютерах с искусственным интеллектом¹. Как сделать механизм с алгоритмами или, например, обучаемой нейросетью, способным поступать нравственно. Пытается дать ответ на вопрос: какой робот этичен?

Она занимается разработкой подходов к установлению нравственных теорий и моделей в устройствах, чтобы создавать искусственных моральных агентов. Она же является ориентиром для разработчиков и инженеров всех предполагаемых искусственных моральных агентов, т.е. имеет междисциплинарный характер развития и ориентируется на разработки в сфере робототехники.

Робоэтика – это область этики, которая концентрирует своё внимание на роботах и существах с ИИ. В этой научной области этики пытаются ответить на дискуссионные вопросы о влиянии робототехники на качество жизни человека, перспективы переосмысления места работа в жизни общества. Ведётся рассуждение о статусе робототехники с точки зрения морали. Эта область имеет много пересечений с юридической и социально-экономической сферой. Самым главным отличием робоэтики от машинной является то, что робоэтика заботится о этичности самого процесса проектирования, строительства, использования и отношения к роботу или существу с ИИ. Тем самым за ней закреплена функция формирования

¹Allen C., Wallach W., Smit I. Why Machine Ethics?// IEEE Intelligent Systems, vol. 21, no. 4, (July-Aug. 2006) pp. 13.

кодексов разработчиков и инженеров робототехники, которые выражаются в единых этических принципах.

Помимо основных проблем, которые разворачиваются внутри этих относительно новых, активно формирующихся областей этического знания, существует другая проблема. «Путаница»\Слияние робоэтики и машинной этикой или отсутствие знания о существовании машинной этики в сознании учёных, инженеров и пользователей робототехники. Такие выводы позволяют сделать результаты работы с научной и публицистической литературой. Англоязычные публикации учёных вводят разделение этих понятия в своих публикациях. В русской же литературе, и научной, и транслируемой СМИ в большинстве случаев происходит подмена машинной этики робоэтикой или их смешивание, таким образом, в обыденном сознании закрепляется смешанный образ этих областей. Усугубляет ситуацию малое количество переводов актуальных теоретических исследований на эту тему на русский язык для закрепления терминологии.

Значение принципов робототехники для машинной этики

В этой работе хотелось бы выразить мысль о том, что робоэтика и машинная этика фокусируются на разных сторонах процесса преобразования общества путём модернизации технологий и внедрения в жизнь общества робототехники, однако они влияют на результаты друг друга.

В 2010 году благодаря работе Совета Инженерных и Физических научных исследований появился набор принципов робототехники². Специалисты, изучающие технологии, право, искусства и социальные науки представили документ, который возник из обсуждения проблем связанных с робототехникой и желания урегулировать процесс в сторону позитивных преобразований в области технологий. Целью создания такого документа стало желание установить план действий по мирной и постепенной интеграции робототехники в жизнь общества с максимальной пользой для

² Principles of robotics//EPSRC URL:

<https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/> (дата обращения: 14.03.18)

каждого человека. Под этим подразумевались работа над доверием к технологии, просчитывание рисков и выгод от коммерциализации и т.д. Все эти действия были вдохновлены известными законом робототехники А. Азимова, их необходимостью, рисками из-за их отсутствия в жизни человека и трудностью воплощения. В этой ситуации иллюстрируется влияние машинной этики и проблемы этического поведения робота на создание этического кодекса для строителей, программистов, инженеров и пользователей роботов, т.е. на область влияния робоэтики.

Роботы относительно малоизвестны рядовым гражданам планеты, как обычное явление их повседневной жизни. Несмотря на это, в краткосрочных и долгосрочных перспективах человечество ждёт активное проникновение робототехники в жизнь людей. Следовательно, задуматься над регулированием деятельности в этой сфере нужно уже сейчас.

На данный момент человеку предлагается руководствоваться правилами из документа. В них принципы представляют два блока сообщений. Первый – базовый и универсальный, второй более специальный и направлен на создание ценностных ориентиров.

Первый блок:

1.Роботы – это многофункциональные инструменты. Роботы не должны быть предназначены исключительно или главным образом для убийства или нанесения вреда людям, кроме как в интересах национальной безопасности.

2.Люди, а не роботы, являются ответственными действующими лицами.

3.Роботы должны быть спроектированы и контролируемы насколько это практически возможно, в соответствии с существующими законами и основными правами и свободами, включая право на неприкосновенность частной жизни.

4.Роботы - это продукты труда (товары). Они должны разрабатываться с использованием процессов, обеспечивающих их безопасность.

5.Роботы - это предметы производства, товары. Они не должны быть разработаны, чтобы эксплуатировать уязвимых пользователей с помощью введения их в заблуждение; вместо этого их машинное происхождение должно быть прозрачным и понятным для всех пользователей.

6.К лицу, несущему юридическую ответственность за робота, следует отнести определённых людей.

Если проанализировать эти принципы с точки машинной этики, то следует указать на пункты 2, 4 и 6 как на решение актуальной проблемы ответственности за деятельность робота и её границ. Принципы переносят её на человека, лишая робота какой-то личной ответственности. При этом самый первый принцип ставит чёткие границы нравственного для роботов с разными функциями, «гражданских и боевых». Запрещая любые виды насилия над человеком, как следствий использования мирных роботов, и закрепляя автономию государства на выпуск роботов, способных причинить вред человеку своей деятельностью.

В дополнение к вышеупомянутым принципам группа также разработала всеобъемлющий набор Сообщений (Посланий). Это второй блок, поддерживающий дух ответственного новаторства. Он призван поощрять ответственность в рамках исследовательского и производящего сообщества робототехники, чтобы установить и укрепить доверие к деятельности этого сообщества.

1.Мы считаем, что роботы могут оказать огромное положительное влияние на общество. Мы хотим поощрять ответственные исследования роботов.

2.Плохая деятельность вредит всем нам.

3.Решение очевидных общественных проблем поможет всем нам добиться прогресса.

4. Важно продемонстрировать, что мы, как создатели роботов, робототехники, привержены исполнять свою деятельность наилучшим образом, согласно стандартам.

5. Чтобы понять контекст и последствия наших исследований, мы должны работать с экспертами из других дисциплин, включая социальные науки, право, философию и искусство.

6. Мы должны учитывать этику гласности и прозрачности: есть ли пределы тому, что должно быть открыто доступно.

7. Когда мы видим ошибочные сообщения в прессе, мы обязуемся уделить время и связаться с журналистами.

Обзор этих принципов более точно раскрывает предмет робоэтики и формирует среду для активного развития машинной этики. Особенно это подчёркивает 5 сообщение, которое призывает сотрудничать участников производства роботов с экспертами из других областей и вместе заботиться о последствиях создания и внедрения новых технологий.

Определение робота

В актуальных проблемах машинной этики фигурирует очень важный элемент (объект), который даёт возможность обособиться новому разделу прикладной и теоретической этики – это робот. Если немного обратиться к истории возникновения этого термина, то обнаружится литературное происхождение этого актуального в научной среде слова. В 1920 году стало возможным попасть в литературный мир известного чешского писателя и драматурга, Карла Чапека. Его научно-фантастическая пьеса «Россумские универсальные роботы» подарила миру неологизм, который прочно закрепился в умах поколений. Благодаря воображению этого человека, на страницах книг появилось общество, где существуют фабрики по производству искусственных людей. Тогда по задумке Чапека их создавали из искусственных материалов, а их предназначением было – выполнять монотонную работу, утомительные, опасные человеческие обязанности за

людей.³ Помимо прямого описания функций роботов, сама этимология слова возвращает читателя к слову «рабство», «раб», которые усиливали низкое положение робота перед человеком. После этого из литературы середины XX века распространилось слово «робототехника», его можно было встретить на страницах произведения Айзека Азимова «Я робот», написанного в 1942 году.

Сейчас на вопрос: «Что такое робот?» люди, которые их создают и люди, которые ими пользуются, чаще всего будут давать разные ответы. Большинство людей ориентируется на образ робота, который транслируется из СМИ, художественных произведений или из их обычной жизни, получается, что современного понимания у человека, который не следит за новостями робототехники, не складывается. Отвечая на вопрос таким образом, что робот – это попытка симитировать человека, только в электронной версии, человек сразу сужает данное понятие и обычный робот-пылесос перестаёт быть роботом. Наоборот, сводя сущность робота к выполнению единообразных простейших функций, за пределами этого понятия окажутся роботы, которые оснащены синтезаторами речи и могут осуществлять коммуникационные действия, быть компаньонами для человека. В связи с тем, что роботы модифицируются с высокой скоростью и для общества не происходит единого схватывания основных технических и функциональных качеств робота, стоит ориентироваться на международные документы. Под роботом в этой работе понимается то же, что вкладывается в понятие робота, разработанное и зафиксированное в международном стандарте по робототехнике: «Робот – это приводной механизм, программируемый по двум и более осям, имеющий некоторую степень автономности, движущийся внутри своей рабочей среды и выполняющий задачи по предназначению».⁴

³ Чапек К. R.U.R. 104 с. URL: <http://lib.ru/SOCFANT/CHAPEK/rur.txt> (дата обращения: 16.03.18)

⁴ ГОСТ Р ИСО 8373-2014 Роботы и робототехнические устройства. Термины и определения: п. 2.6 URL: <http://docs.cntd.ru/document/1200118297> (дата обращения: 10.02.18)

Выделяется четыре основных свойства, присущих роботу⁵:

Первое свойство обеспечивает ему возможность передвигаться, действовать в самых разных условиях, будь то открытая местность, больница или офис – это мобильность.

Второе свойство стало возможным благодаря датчикам и приводам, которые собирают нужную (релевантную) информацию от внешней среды и позволяют роботу адаптироваться и действовать в условиях окружающих условий – это интерактивность.

Коммуникация – третье свойство робота, которое осуществляется, благодаря компьютерным интерфейсам, системам распознавания голоса и синтеза речи.

Помимо этого активно обсуждается и старается реализоваться способность к автономии робота, которая выражается в том, чтобы «думать» самостоятельно и выносить собственные решения, тем самым воздействуя на внешнее окружение без прямого внешнего контроля со стороны человека.

Современная робототехника, как правило, оснащается различными формами искусственного интеллекта (ИИ). Копирование человеческого сознания, интеллекта и внедрение их к компьютерным системам в результате позволяет машинам действовать в силу определённого уровня интеллекта, например, воспринимать и осознавать изменения в окружающей среде и в соответствии с этим планировать своё функционирование. Для последнего качества роботов, из тех, что были перечислены выше : автономности, ИИ играет ключевое значение, поскольку он позволяет роботам выполнять сложные задачи в различных средах, подвергающихся постоянным изменениям своей структуры, таких как вождение автомобиля на оживлённой трассе и адаптация на дороге при полной автономии, независимости от человеческого руководства.

⁵ Report Of Comest On Robotics Ethics//UNESCO; 2017 pp. 4. URL: <http://unesdoc.unesco.org/images/0025/002539/253952E.pdf> (дата обращения: 10.03.18)

Выполнять свои задачи роботам помогают алгоритмы, особые правила или инструкции для решения задач. Доминируют два типа алгоритмов для роботизированных устройств. Первые, детерминированные алгоритмы, которые контролируют предсказуемое поведение детерминированных роботов. Второй вид – это ИИ или алгоритмы стохастические, вероятностные, со способностью к обучению. Они формируют «сердцевину» (центр) когнитивных роботов. Детерминированное поведение робота, даже при всех его автономных и сложных свойствах, запрограммировано и определено по существу. Однако с (умственными) когнитивными роботами дело обстоит многим сложнее и интереснее для учёных и людей в целом. Такие роботы обладают способностью обучаться на прошлом опыте своей деятельности и выверять свои собственные алгоритмы для деятельности. Это говорит о том, что нельзя утверждать о полной предсказуемости их поведения для человека, что заставляет задавать много вопросов этического характера.

В Толковом словаре искусственного интеллекта выделяют три вида роботов⁶.

Первый тип - это автономный робот. Его характерными функциями являются наличие информации об окружающей реальности и её особенностях, а так же присутствие программы, направленной на анализ среды, нахождение ситуативных решений и накопление необходимых знаний для деятельности в зависимости от места нахождения и характера ситуации. Автономный робот – это интеллектуальная система: «Техническое устройство, способное к планированию целесообразного поведения в условиях динамической, заранее не полностью известной среды».

Второй тип робота отражает некоторые способности человеческого тела: зрение и хватательные движения рук. Интегральный робот – это действующее существо, относимое к интеллектуальным системам, которое способно к координированию происходящего. Система «глаз-рука» работает

⁶ Аверкин А.Н., Гаазе-Рапопорт М.Г., Поспелов Д.А. Толковый словарь по искусственному интеллекту URL: <http://www.raai.org/library/tolk/aivoc.html#L489> (дата обращения: 25.04.18)

со средствами перемещения и с движениями манипуляторов. При этом такой робот должен проводить аналитическую работу связанную с отображением реальности и уметь принимать решения , беря за основу анализ данных зрительных сцен.

Третий тип роботов оснащён всеми техническими компонентами, свойственными интеллектуальным системам. Функционал такого робота предполагает накопление знаний об окружающей реальности, на их основе программируются планы для выполнения поставленных задач и целей. При этом отличительными функциями являются – общение с внешними агентами взаимодействий и построение программ поведения.

Подводя итог, можно выделить базовые компоненты устройства любого робота – это его электромеханическое устройство, способность вступать во взаимодействие с окружающей средой, мобильность устройства, степень свободы, программируемость, управляемость или автономность.

Сильный и слабый искусственный интеллект

Представление о ИИ очень важно для формирования более полного и цельного представления о проблемах машинной этики, поэтому некоторая часть исследования была посвящена поиску определения ИИ и определению положения ИИ в гуманитарной и инженерной среде.

Рассматривая ИИ как часть современной культуры, специалисты гуманитарной сферы сформировали несколько разделов, которые символизируют ИИ. В качестве ИИ могут выступать экспертные системы, робототехника, нейронные сети. При этом ИИ – это не только конкретное действующее лицо, но умения, навыки, способности – среди которых выделяют взаимодействие ЭВМ и человека, автоматическое программирование, когнитивное моделирование, обучение, поиск, управление и планирование, способность зрительного различения изображений и их обработки. Естественный язык, логическое программирование наряду с другими понятиями логики тоже входит в понятие ИИ. Применение ИИ – это тоже так называемый ИИ.

В системе понимания ИИ – специалисты фокусируются не на одном определении и способности ИИ. Ю.Ю. Петрунин считает, что люди склонные выделять несколько характерных определений искусственного интеллекта, закрепившихся в системе знаний об этом феномене.

Первые, по мнению Ю.Ю. Петрунина склонны делать вывод, что ИИ – это прежде всего умение решать сложные, подвластные человеческому интеллекту, задачи. Вторые устанавливают для себя, что ИИ – это способность обучаться, находить обобщения и аналогии. Последние замкнули понятие ИИ на том, что это возможность взаимодействия с окружающим миром с помощью коммуникации, восприятия и осознания полученной информации. В итоге ИИ имеет большой список характерных черт. В данный момент более 100 факторов могут обеспечивать интеллектуальную деятельность, половина из которых изучена учёными, поэтому вытекает такое разнообразие пониманий и представлений о ИИ.

Среди инженеров и учёных в сфере компьютерных технологий существует два определения для ИИ. Первое представляет ИИ, как направление научной деятельности «в рамках которого ставятся и решаются задачи аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными». Второе определение описывает способность «интеллектуальных систем выполнять функции (творческие), которые традиционно считаются прерогативой человека».⁷

При этом в философии сформировалась дискуссионная проблема, связанная с сильным и слабым ИИ. Считается, что сильный искусственный интеллект будет сопоставим с человеческим по уровню способностей и возможностей. Однако на данный момент, технологически человечество не имеет возможности сконструировать интеллектуальную систему таких мощностей. Для того чтобы в один из моментов прогресса в робототехнике у

⁷Аверкин А.Н., Гаазе-Рапопорт М.Г., Поспелов Д.А. Толковый словарь по искусственному интеллекту URL: <http://www.raai.org/library/tolk/aivoc.html#L208> (дата обращения: 27.04.18)

учёных была возможность проверить способности интеллектуальной системы существует тест Тьюринга, который считается тестом для проверки способностей ИИ соответствовать человеческому интеллекту так, чтобы человек не смог отличить его от настоящего. Это эмпирический тест, идея которого была предложена Аланом Тьюрингом в статье «Вычислительные машины и разум», опубликованной в 1950 году в философском журнале Mind. При «слепом» взаимодействии, переписке, человека-судьи с другим человеком и интеллектуальной системой (ИС), человек должен отличить с кем он взаимодействует, если условный робот будет воспринят судьёй как человек, то тест считается пройденным успешно.

Только существование теста Тьюринга не гарантирует установление за ИС статуса сильного ИИ. Этому мешает логический эксперимент под названием «Китайская комната», который провёл известный логик Джон Сёрль. Благодаря его эксперименту стали ставиться под сомнения критерии оценивания сильного ИИ, понимания мышления в научном обществе и существования сильного ИИ в целом. Тем самым оставляя для сильного ИИ место только в фантастической литературе.

Слабый ИИ во многом уступает человеческому, при этом споры о его существовании не ведутся. Примерами слабого ИИ могут считаться нейросети, например разработка компании «Яндекс»- «Алиса» и «Google» - «Siri», так же это различные системы по типу «Умный дом» и «Умный город», которые уже сейчас меняют облик реальности и становятся неотъемлемой частью настоящего и будущего человека.

Подводя итоги этой части работы, нужно отметить, что удалось дать определение машинной этике, как направлению этической мысли, которое находится в поиске подходов, концепций и решений для успешного воплощения механизма принятия этических решений в системах и роботах, обладающих ИИ. Это направление является одним из составляющих большой области этики ИИ, так же как и робоэтика. Робоэтика и машинная этика концентрируются на проблемах использования интеллектуальных

машин, но смотрят на это под разным углом. Робоэтика задаётся вопросом, как нам следует вести себя и относиться к роботам, а машинная этика разрабатывает подходы того, как машина может относиться и вести с себя при социальном взаимодействии с человеком. Однако знания, которые добываются и в робоэтике, и в машинной этике взаимосвязаны и для реализации принципов и подходов обеих необходимо учитывать всю накопленную информацию. Зная о том, что в науке понимают под роботом и поняв, что ИИ сегодня представлен больше слабой версией, которая не несёт в себе угроз словно фантастическая литература, можно продолжить рассматривать какие проблемы в теоретическом и прикладном срезе формируют машинную этику.

Глава 2 Теоретические проблемы машинной этики

Многие уверены, что машинная этика – это ключ к работе с новыми технологиями и успешному внедрению этих технологий в жизнь человека. Однако на вопрос о развитии этой сферы существует не только позитивный отклик. Некоторые люди, рассуждая о машинной этике, склонны обращаться к идее, что без эмоционального подкрепления деятельности роботов и интеллектуальных систем, нельзя говорить о машинной этике. Это заявление могло бы обратить нас к другим теоретическим проблемам этики о источниках морали и условиях для их возникновения. Ясность имеет лишь одно утверждение, что расширение и развитие индустрии робототехники приведёт к тому, что учёные и инженеры обратятся к мыслям, которые развивает машинная этика.

Утверждая насущную необходимость машинной этики для общества и современной научной среды, зная её определение, можно перейти к актуальным проблемам, которыми она занимается. В этой главе будут рассматриваться теоретические проблемы, касающиеся машинной этики.

Искусственный моральный агент

При изучении научной и публицистической литературы может показаться, что теоретические вопросы не так популярны, как этические проблемы прикладного характера. Применения искусственных моральных агентов (ИМА) и их взаимодействия с человеком в определённой среде является интересной темой для рассуждений и исследований, само общество заинтересовано в прогнозах и исследованиях такого типа. В это же время существует трудность определения морального агента, это вопрос теоретического характера машинной этики.

Джеймс Мур в своих исследованиях выделяет 4 вида моральных агентов, рассуждая о машинах как возможных моральных агентах⁸.

⁸ Moor J.H. The Nature, Importance, and Difficulty of Machine Ethics//IEEE Intelligent Systems, Vol. 21, No. 4 (2006) 19.

Первая группа моральных агентов - это машины, которые являются агентами этического воздействия.

Среди положительных примеров того, как робот может воздействовать на человека, можно продемонстрировать давний опыт Катара⁹. В Катаре большой популярностью пользуются традиционные верблюжьи бега, выигрыш в таком забеге может принести победителю огромную денежную сумму. Основным гарантом успеха при отличном состоянии верблюда в таких забегах является компактный жокей с маленьким весом. Наилучший жокей в таком случае - это ребёнок. Управление верблюдом в раннем возрасте очень опасно, поэтому обладатели животных покупали детей, брали их в рабство, учили их управляться с животным, держали без образования и в голоде, потому что вес жокея – очень важный показатель. Когда дети вырастали и больше не могли заниматься гонками, их бросали и не заботились об их дальнейшей судьбе. В обществе и государстве вопрос о приемлемости гонок с участием невольных детей оставался дискуссионным. Переломным моментом в этой ситуации стала смерть одного из жокеев, после чего государство не могло игнорировать проблему. Решением стал запрет на использование детей младше 18 лет, однако традиция бегов осталась. Позже детей было предложено заменить на роботов, т.к. взрослые наездники не устраивали владельцев верблюдов. В 2003 году началась разработка робота «Camel» компанией из Швейцарии. В 2006 этот робот весом в 22 кг держал победу в гонке среди жокеев-людей и после этого роботов стали активно использовать в бегах. Тем самым государство сохранило культурные традиции бедуинов и решило проблему рабства. В этой ситуации робот стал решением вопроса и способствовал укреплению ценности свободы человека, при этом трансформировал культурную традицию народа в этически приемлемую практику.

⁹ Теплов О. Железные всадники // GEO 2006 URL: <http://guide.travel.ru/qatar/124106.html> (дата обращения: 08.05.2018)

Среди негативных примеров можно представить системы-вредители в сети, которые заражают вирусами устройства граждан и устройства, связанные с работой государства. На данный момент конкретных примеров негативного влияния нейросетей или роботов с ИИ нет, потому что разработчики наоборот развивают нейросети, как инструмент по борьбе с киберпреступлениями. Однако кинематограф прекрасно предположил такую ситуацию в фильме «Терминатор. Восстание машин», где причиной ядерной катастрофы стала программа с ИИ «Skynet», которая создала вирус, взломавший компьютерную систему Пентагона.

Второй тип агентов – это неявные этические агенты. Их главная характеристика - совершение морального действия по принципу избегания неэтичных решений. Неявные моральные агенты содержат программное обеспечение, которое поддерживает этическое поведение. Программа в теории работает таким образом, что внутренние функции будут поддерживать этически приемлемое поведение и избегать действий, которые порицаются в обществе. В этой ситуации – можно говорить о «природной» нравственности машины, которая имеет свои достоинства, хоть и не является идеальной.

Примерами неявных моральных агентов являются компьютеры, которые по своим функциям и задачами обращены к вопросам безопасности или достоверности. Системы банковских переводов, автоматические кассовые аппараты – представители этого типа морального действия. Ведь транзакции должны проходить максимально точно, а машина должна быть сконструирована таким образом, чтобы выдавать определённую сумму денег при проведении банковской операции.

Так же автопилотируемый самолёт является примером неявного этического агента. Безопасное прибытие самолёта в нужное время в пункт назначения обеспечивается программой автопилота, которую разработали инженеры. Другие системы самолёта держат в курсе автопилот о близко находящихся объектах, о количестве и запасах топлива – это этические итоги

деятельности инженерной команды, проектирующей действия автопилота. Так же ярким примером являются программные обеспечения аптек, которые проверяют наличие лекарств, их взаимодействие с другими препаратами, помогают фармацевту выписать лекарства, которые принесут пациенту больше пользы, чем вреда.

При всех достоинствах этих машин, они не никак не дают возможности говорить о них, как о полных моральных агентах. Так как их действия основаны на алгоритмах и не предполагают той степени свободы выбора, которой обладает человек при совершении морального действия.

Третий тип моральных агентов – это явные моральные агенты. Так как явное моральное действие машины является неуловимым в практике процессом на данный момент, для теоретиков – это очень актуальный предмет для размышлений. Под явным этическим действием предполагается представление и понимание этических категорий, анализ поступков с помощью понятий «хорошего» и «плохого» и других этических категориях, представление об этике и деятельность на основе этих знаний. Все эти элементы должны выполняются машиной, наделённой ИИ, для того, чтобы называть её явным моральным агентом.

Учёные проводят работу с этическими теориями, проводят эксперименты с роботами¹⁰. Робот Джереми, названный в честь И. Бентама, реализует анализ вероятности удовольствия или страдания лица, при совершении какого-либо действия. Джереми ориентирован на использование утилитарной этической теории и этические выводы будет выносить исходя из вероятного количества перенесённого человеком счастья и страдания.

Другой робот, названный в честь Дэвида Росса, опирается на его теорию правильного действия, которое получается при соотнесении фактического обязательства и обязательства *prima facie* при участии моральной интуиции агента. В такой системе нет ранжирования и

¹⁰ Anderson M., Anderson S.L., and C. Armen Towards Machine Ethics: Implementing Two Action-Based Ethical Theories//Machine Ethics. AAAI Press, (2005) . pp. 3.

приоритетности обязательств, а моральной интуиции у роботов нет, поэтому возникает трудность в создании алгоритма на основании этой теории.

Андерсоны и Армин, работающие над этими роботами, разработали компьютерную модель, которая преодолевает неопределённость и неясность между обязанностями робота, которые бывают выражены как фактические предписания и обязанности *prima facie*¹¹. Модель использует алгоритм обучения для того, чтобы корректировать суждения об обязательствах, принимая во внимание и фактические обязательства, и прошлые интуиции о похожих и непохожих случаях, связанных с этими ситуациями. Все эти модели задают продуктивную тенденцию в области исследования алгоритмов в ключе разных этических теорий и создания явного морального агента, которого на данный момент не существует.

Последний, известный на данный момент тип, это полноценный моральный агент. Полноценным может считаться моральный агент, который в полной мере реализует нравственную деятельность, действует исходя из этических соображений, имеет способность к моральной оценке. Здесь говорится о машинах, которые будут способны сравнить свои способности к моральной деятельности с человеческими. Люди выступают полноценными моральными агентами, обладающими сознанием, целеполаганием и свободной волей.

Сопоставляя машину и человека в таком ракурсе, мнения разделяются. Некоторые считают, что между человеком и машиной стоит яркая и непреодолимая грань, через которую машине никогда не переступить. Это онтологическое разделение субъектов, человека и машины, имеет несколько видов аргумента против искусственных полных моральных агентов.

Во-первых, можно утверждать, что моральным агентом считается только полноценный моральный агент, т.е. на данный момент только человек. Тем самым нивелируется этическая составляющая деятельности остальных

¹¹ Anderson M., Anderson S.L., and C. Armen Towards Machine Ethics: Implementing Two Action-Based Ethical Theories//Machine Ethics. AAAI Press, (2005) . pp. 5.

трёх типов агентов, о которых писалось ранее. Утверждая это, машинная этика лишается моральных агентов вовсе. Те, кто выступают против этого аргумента, считают, что значение этики в машинах, которые подходят под типы явных и неявных моральных агентов и агентов этического воздействия не так явно и ярко выражено, это проявление не должно оставаться проигнорированным. Ведь изучение и внедрение этического компонента напрямую зависит от разработчиков, живых людей.

Другой формой аргумента против возможности машин быть полноценными этическими агентами выражается в том, что ни одна машин не сможет стать таким агентом. Машине не свойственна свободная воля, сознание и целеполагание по определению.

В данном исследовании, решающим и наиболее близким для автора является аргумент против этого заявления, который выражается в неопределённом исходе итогов развития индустрии. Научное общество не может быть уверено в том, что машина не может стать полноценным моральным агентом. Нельзя сказать, что на этом пути робототехника стала в тупик. Она лишь на пороге своего развития и возможно, машины будут наделены такими качествами, а возможно нет. В этом споре нельзя поставить точку в данное время.

В случае машин можно говорить о том, что действующее лицо выступает как нормативный агент, если алгоритмы, описывающие и задающие действие, не способны к моральной оценке мотивов и результатов предполагаемого действия. Итогом по этой теме станет определение морального агента, сделанное на основе описанных данных. Действующее лицо, способное выражать моральную оценку, опирающееся на моральные нормы, образцы и этические категории в своей деятельности можно назвать моральным агентом.

Мораль искусственного интеллекта

Главная цель этой области исследований – создание искусственных агентов, которые будут действовать, словно моральные агенты. Эта цель является основной.

Помимо неё есть промежуточная цель, которая объясняется актуальным состоянием и развитием индустрии робототехники. Промежуточная цель несёт в себе задачу приобщить и сделать максимально чувствительными системы с ИИ к ценности, этичности и легальности деятельности как таковой.

Полноценная активность в разработке искусственной морали обеспечивается с помощью изучения одновременно и технологических вопросов, и философских вопросов, которые связаны с созданием морального агента в лице интерфейса робота или программы, использующей ИИ.

Искусственная мораль позволяет в перспективе перенести часть ответственности за поведение искусственного морального агента с разработчиков (программного обеспечения) ПО и инженеров на самих агентов с ИИ. Задача сделать агентов с ИИ – моральными становится всё актуальнее, потому что автономность агентов становится всё шире из года в год, благодаря прогрессу в сфере робототехники. Их способности уже сейчас не всегда возможно оценить с точки зрения этичности в реальном времени из-за скорости и сложности выполняемых ими операций.

Программирование возможности совершать моральный выбор агентам с ИИ – выбирать уместен ли тот или иной поступок с точки зрения этики, предлагается для тех агентов, которые являются «чувствительными» к этим понятиям. При этом эта способность не должна снимать ответственность с тех, кто пользуется или создаёт этих агентов. Что бы ни лежало в основе морального выбора искусственного агента, подлинная мораль или искусственная, этот выбор несёт определённые результаты и последствия. В отличие от случайного вреда какой-либо вещи, вред, принесённый

искусственным агентом можно контролировать и регулировать со стороны этих агентов.

К задаче создания искусственных моральных агентов (ИМА) возможно подойти с точки зрения разных подходов. Первый из них является контролирующим поведение с помощью разработанных правил и не требует вмешательства в механизм машины. Второй подход является более сложным для разработчиков и инженеров, потому что требует создания чувствительной системы, способной обучаться и формировать алгоритмы приемлемого нравственного поведения.

Директивный подход создания искусственной морали.

ИМА – могут использовать моральные принципы или теории в качестве правил для выбора этически приемлемого поведения в обществе. В этом заключается смысл директивного подхода к созданию искусственного морального агента¹².

Директивный подход имеет стороны, которые в первую очередь подвергаются критическому осмыслению. Первая претензия к осуществлению этого подхода - невозможности обеспечить главной, доминирующей теорией ИМА для выполнения поступков. Этот подход теоретически проявил ненадёжность для множества задач, которые могут совершать (искусственные агенты) ИА. При том, что параллельно с этическими исследованиями, директивный подход используется в других сферах робототехники и проявляет себя более успешно и обеспечивает наилучшее технологическое решение задач ИА-ом.

Несмотря на подмеченные слабые стороны этого подхода, Аллен, Смит и Валлах в своей работе отмечают, что перспективы и вопросы, сопутствующие этому подходу необходимо внимательно рассматривать.

¹²Wallach.W., Allen C. Smit I. Artificial morality: Top-down, bottom-up, and hybrid approaches//Ethics and Information Technology (2006) pp.149.

Наиболее явно возражения против этого подхода можно заметить только при рассмотрении перспектив встроенных в ИМА процедур принятия решения, которые основаны на внешних моральных теориях.

Принципы, как кандидаты для перехода к алгоритмическим процедурам принятий решений, варьируются от религиозных идеалов и моральных кодексов, до культурных ценностей и философских систем. Золотое правило нравственности, 10 Заповедей, утилитаризм или кантианское долженствование – это лишь некоторые из возможных источников правил для директивных этических систем. 4 закона робототехники, которые описал Айзек Азимов так же можно отнести к разновидности директивной системы. Отличие понимания той или иной ценности в разных этических системах уже может породить споры о способах реализации директивной этической системы ИМА.

Гипотетически считается, что директивная система представляет набор правил, которому должно следовать моральному агенту. При одном из осмыслений можно представить эту систему в виде разнородных правил того, что запрещено делать и того, что следует.

Заповедность морали, уходящая истоками в иудейскую традицию, повторяется и в законах робототехники Азимова, считают Аллен, Смит и Валлах¹³. Для одного вида поведения существует одно правило. При этом авторы видят проблему в модели правил по типу заповедей в том, что они часто вступают в противоречие друг с другом. С точки зрения вычисления, подсчёта ситуации – такие конфликты трудноразрешимы, если нет более авторитетного принципа или правила, которое позволит разрешить конфликт этих правил. Законы А. Азимова представлены в виде 4 правил:

1. «Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.

¹³Wallach W., Allen C. Smit I. Artificial morality: Top-down, bottom-up, and hybrid approaches//Ethics and Information Technology (2006) pp.150.

2. Робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Законам»¹⁴
4. Нулевой закон: «Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинён вред»¹⁵.

Большинство директивных систем не дают ответа на вопрос, как разрешить такие разногласия. А. Азимов попытался разрешить эту проблему таким образом, что предыдущий закон авторитетнее последующего, т.е. робот, ориентируясь на второй закон и найдя противоречие, может сослаться на первый закон. Для этого был разработан и нулевой или четвёртый закон робототехники, который направлен на защиту всего человечества (надо процитировать).

Конфликт правил можно заметить между первыми двумя правилами, а существование нулевого правила не даёт решения, не описывает понятие вреда и не разрешает дилемму в случае несовместимости обязанностей: кому нанести вред из двоих, если этого нельзя избежать или кому причинить вред, если, спасая одного, навредишь другому?

На замену конкретным правилам для поступков ИМА, возможно сделать упор на следование более абстрактным универсальным принципам. Они стали бы основой для выведения точных решений для каждой отдельной ситуации. В этики такие системы принципов представляет, например Кантианское учение о долженствовании и утилитаристское учение. Обе теоретические модели могут способствовать созданию ИМА, этических роботов.

Утилитаризм очень привлекателен возможностью и необходимостью вычислений страданий и удовольствий, которые может причинить ИМА. К

¹⁴ Азимов А.Я, робот Эксмо; 2010 стр. 392

¹⁵ Азимов А.Я, робот Эксмо; 2010 стр.556

сожалению, его сила является и непреодолимой проблемой несоизмеримости некоторых удовольствий и страданий. При попытках создать эквивалент для измерения счастья и горя всё терпит неудачи, потому что трудно сказать на что человек мог бы выменять свои радости и что помогло бы ему откупиться от чьего-то горя. При всех предположениях решения проблемы несоизмеримости, следующим препятствием станет проведение исчислений и калькуляций общего блага и вреда «сверху», все подобные исчисления будет необходимо проверить, просчитать альтернативы действий и их результаты, при этом необходимо учитывать множество мнений.

Напротив, кантианская модель, ориентируется не на результат, а на мотив поступка и предъявляет требования к действующему лицу уважать конкретные права и обязанности каждого человека. Для решения проблемы противоречащих друг другу конкретных правил и принципов существует императивная форма и категорический императив служит её примером. Для воплощения подхода и создания алгоритма для ИМА, так же как и в случае утилитаризма, необходимо совершить много вычислений, чтобы добиться исчерпывающей моральной оценки любого поступка, которыми будет руководствоваться ИМА. Для этого необходимо быть уверенным не только в истинных мотивах действующего лица, но и характеристике мотивов и их оценке. Для этого надо удостовериться, как мотив воздействует на рационального субъекта и возможны ли какие-то отхождения от единой схемы и выполнение другого действия, исходя из одного мотива. В этом случае учёным и философам обязательно быть высокими специалистами в области психологии поведения, этике и разобрать все моральные действия в мире. На данный момент – это лишь можно представить, но реализовать маловероятно.

Можно сказать, что консеквенционалистские и деонтологические подходы сталкиваются со своими уникальными проблемами исчислений. В это время в этих подходах возникает одинаковый вопрос, который касается возможности сбора и сравнения собранной информации человеком, ИМА

или компьютером, которая будет нужна для применения в определённое время.

Восходящий подход к установке искусственной морали

Подход, который принято называть «восходящим» подразумевает развитие ИМА без навязывания определённой моральной теории, а с помощью установления подходящей среды¹⁶. В этой среде ИМА проходит своё развитие через систему воспитания путём прививания правильного поведения через систему награждений. Соответственно плохие поступки считается возможным искоренить благодаря наказаниям. Процесс моральной «социализации», обретения нравственной восприимчивости предполагает комплексный подход. Обучение через опыт предполагает развитие путём проб и ошибок бессознательной части механизма, для которой придумывают новые трудности и задания программисты и инженеры, либо это воспитательное развитие обучаемого механизма.

Каждый из методов имеет характеристики схожие с тем, как маленький ребёнок получает нравственное воспитание в социальной среде, в которой выявляется уместность того или иного поступка и поведения в целом. Главное, что чёткой теории о принципах хорошего и плохого поведения для ребёнка, а в случае исследования для ИМА, никто не приводит.

Восходящие стратегии ставят целью привить навыки и стандарты, которые являются неотъемлемой частью общей структуры системы, но их чрезвычайно трудно развивать. Развитие и обучение представляют собой метод проб и ошибок – обучение через неудачи. Эта работа будет занимать много времени, несмотря на высокие скорости мира компьютерной обработки информации и работы эволюционирующих алгоритмов.

Предполагается, что такое обучение должно иметь свои рамки, курс, цели и задачи, что превращает его в определённый образовательный режим, который включает нравственное воспитание. Имитация детского разума и

¹⁶Wallach.W., Allen C. Smit I. Artificial morality: Top-down, bottom-up, and hybrid approaches//Ethics and Information Technology (2006) pp.151.

сознания у механизма – это лишь один вариант, используемый для создания интеллектуальных агентов, способных к обучению.

Что считается важным компонентом для воспитательного режима ИМА – это моделирование развития, проектирование вычислительных обучающих платформ и ассоциативные методы обучения.

На этом моменте стоит сделать заключительные выводы, которые касаются теоретических проблем машинной. Первой проблемой становится определение искусственного морального агента – это действующее лицо, способное выражать моральную оценку, опирающееся на моральные нормы, образцы и этические категории в своей деятельности. На данный момент выделяют 4 типа моральных агентов: первые из них самые простые и лишь могут влиять на поступки человека, изменяя и формируя особые этические границы реальности, моральное сознание, в которых совершается поступок. Следующие два типа агентов с помощью алгоритмов могут совершать этические действия, но агент второго типа будет действовать по принципу избегания плохих поступков, а агент третьего типа будет стремиться делать хорошие поступки. Последний тип моральных агентов представлен человеком и подразумевает полную, автономную деятельность с пониманием и осознанием этического контекста реальности: норм, ценностей, идеалов поведения и т.д. Самое главное, что в попытках создания морального агента нельзя спускать только на уровень алгоритмов и логики, потому что в этом случае может возникнуть только нормативный агент, но ещё не моральный.

Второй основной проблемой является создание отлаженного механизма для машин, чтобы реализовать способность выносить этические суждения и совершать поступки, исходя и из этического контекста. На данный момент источниками искусственной морали предлагают быть установленные правила, которые словно программа будут вести отчёт о деятельности. Во втором случае машины предлагают оснастить системами, способными усваивать хорошее, приемлемое поведение в обществе. Помимо разницы подходов существует вопрос того на какие источники нравственности

человек будет опираться при составлении программ для машин, на эти роли претендуют и деонтологические, и консеквенциалистские теории, а так же теории добродетелей. Все они предлагают свои механизмы принятия решений и по-разному определяют значение мотива и результата деятельности, особенно это важно для последующей оценки поступка. На данный момент нет единого решения, все разработки требуют тщательных и многолетних испытаний с участием людей.

Глава 3 Прикладные проблемы машинной этики

В этой главе будут описаны проблемы, с которыми на данный момент столкнулся человек при использовании машин с ИИ. Помимо насущных вопросов будут рассмотрены дискуссионные вопросы, касающиеся прогнозов использования уже существующих машин с ИИ в обозримом будущем.

Для рассмотрения проблем были взяты только две отрасли использования машин с ИИ – военная и социальная. Военная область применения исходит из социального взаимодействия робота с человеком, так же как и социальная. Однако активное использование роботов в военных действиях не исключает наличие насильственной, деструктивной деятельности робота с врагом. Этот аспект взаимодействия человека и машины с ИИ является остро обсуждаемым.

Социальная сфера будет представлена роботами в медицине, терапии и хирургии, и социальном уходе за различными слоями людей. Решение остановиться на нескольких областях использования робототехники исходило из желания дать более подробный анализ проблем, привести больше примеров и предложить возможные решения.

Понятие и виды боевого робота

В первую очередь нужно обозначить характерные особенности боевых роботов, которые присущи именно им. Боевой робот отличается от обычного робота, понятие которого было рассмотрено в первой главе тем, что механизмы используются в особой среде ведения военных действий. Они проектируются в зависимости от целей и задач, которые им предстоит выполнять. Такие устройства создаются в большей степени для того, чтобы заменить человека в боевой ситуации, тем самым сохраняя ему жизнь¹⁷. Так же военные роботы создаются для работы в опасных для человека условиях или условиях несовместимых с его возможностями поддерживать жизнь.

¹⁷Коваливский А. Боевые роботы. Конкуренция за будущее //Новый оборонный заказ .Стратегии №5(42) 2016. URL: <http://dfnc.ru/robot/boevye-roboty-konkurentsya-za-budushhee/> (дата обращения: 09.04.18)

В первую очередь людям при упоминании боевых роботов приходят в голову образы антропоморфных механизмов, которые обладают полной автономностью и служащие для истребления военных и гражданских людей при ведении военных действий. Однако военная робототехника в настоящее время работает, как устройство телеприсутствия, позволяющее человеку удалённо наблюдать за ситуацией и иметь возможность вмешаться и повлиять на её развитие. Лишь в некоторых случаях можно говорить о полуавтономных или автономных боевых устройствах. Так же много боевой робототехники не имеет цели поражать противника, а выполняет очень много других задач.

В зависимости от рода войск выделяют военных сухопутных, морских и воздушных роботов.

Роботы сухопутных войск выступают в качестве сапёров, санитаров, эвакуаторов и тягачей. Так же можно представить беспилотные автомобили, тактические машины. Помимо этого под сухопутными роботами понимают мобильные робототехнические комплексы, которые занимаются наблюдением, разведкой, обезвреживают взрывоопасные предметы. Их разработкой занимаются и российские военные инженеры. Среди самых известных комплексов в РФ выделяют управляемый комплекс «Нерехта», так же робототехнический комплекс «Уран-9», который способен вести полноценный бой, и МРК – мобильный робототехнический комплекс «Арго» и «Платформа-М». Так же в 2017 году был представлен новейший комплекс «Нахлебник» от концерна «Калашников». Среди сухопутной робототехники в мировой военной практике популяризирован метод дистанционного управления, лишь отдельные экземпляры оснащают ИИ. Это вызвано большими затратами и сложностью воплощения.

Роботы морского флота бывают подводными и надводными. Одни из известных моделей - «Гном», REMUS и Transphibian. Первый является телеуправляемым, второй дистанционно управляется двумя операторами, а последний единственный автономный робот. Для роботов морских войск

главными и возможными для выполнения задачами являются операции по патрулированию водных территорий, охране береговой линии, поиску мин. Так же роботы могут быть направлены на выполнение поисково-спасательных работ и исследования потенциально опасных объектов. В целом, сухопутная робототехника более специализируется на дистанционных методах управления, а морская робототехника в целом не так разнообразна и автопилотируемые системы в ней распространены несильно.

Гораздо обильнее автопилотируемые роботы представлены на вооружении военно-воздушных сил. БПЛА или беспилотный летательный аппарат определяется как «летательный аппарат, выполняющий полет без пилота (экипажа) на борту и управляемый в полете автоматически, оператором с пункта управления или сочетанием указанных способов»¹⁸. Именно БПЛА представлены как дистанционно-пилотируемые и полностью беспилотные механизмы. Помимо военных целей БПЛА могут использовать для картографирования местностей или для сельскохозяйственных нужд. Мировыми лидерами по производству БПЛА является Израиль, США, Канада и КНР. Однако на их счету не так много ударных беспилотников, больший упор делается на мирные, обслуживающие БПЛА. Ударные БПЛА способны нести боевую нагрузку и их целью могут быть уничтожение воздушных, наземных и морских объектов. MQ-1 Predator и MQ-9 Reaper – это наиболее распространённые БПЛА. Они показали мощности и качество, участвуя в военных операциях США в Ираке и Афганистане.

Именно так выглядит современное состояние военной техники, оснащённой ИИ, среди которых получили наибольшее применение в военных действиях именно БПЛА.

Боевые роботы и проблемы машинной этики

Среди всех упомянутых видов роботизированной боевой техники излюбленным предметом для этических рассуждений и кейсов являются

¹⁸Постановление Правительства РФ от 11.03.2010 N 138 (ред. от 12.07.2016) «Об утверждении Федеральных правил использования воздушного пространства Российской Федерации» URL: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=201851#0> (дата обращения: 04.08.18)

БПЛА. В первую очередь они интересны тем, что их автономность подтверждена их непосредственным использованием. Они отражают все страхи и прогнозы, которые выстраиваются вокруг военной робототехники. Научные базы полны статей и исследований, связанных с беспилотными ударными системами, СМИ популяризируют проблемы внедрения роботов на военное поприще, однако тщательный анализ актуальных этических проблем, которые вызывают научный интерес и обыденное любопытство, значительно сокращает круг проблем военной робототехники в области машинной этики. Действительно большинство вопросов, касающихся доверия командующего к роботу исполнителю, проблемы ответственности за боевые преступления или вопрос глобальной стратегической стабильности в реальности, которая заполнена боевой робототехникой относятся к области робоэтики, которая, как уже установилось ранее имеет регулирующую функцию по отношению к создателям и пользователям машин, оснащённых ИИ.

Одной из первых проблем выступает право на убийство человека роботом. Если отбросить юридическую грань вопроса о правовом статусе можно посмотреть, чем может руководствоваться робот, как моральный агент, который создан для замещения военных, на поле боя. В этом случае на первое место выходят пути решения, которые представляет искусственная мораль.

Во-первых, военный робот может пользоваться этической теорией с присущими ей принципами, чтобы избрать верное действие. Это могут быть теории добродетелей, консеквенционалистские, деонтологические теории и т.д. Во-вторых, робот может быть ограничен в принятии решений правилами алгоритма, которые нельзя нарушать, ими могут стать законы подобные 4 законам робототехники А. Азимова. Об особенностях тех или иных подходов уже было сказано в прошлой главе исследования.

На данный момент эта проблема находит выход в обращении к принципам робототехники, которые были раскрыты в первой части

исследования. Именно тот принцип, где роботы с ИИ, интеллектуальные системы не обладают правовым статусом, они имеют статус продукта и ответственность за их деятельность несут разработчики и пользователи, поэтому вопрос ответственности за причинение вреда, насильственные действия, убийства лежит на тех, кто изготавливает и использует машины в таких целях.

Помимо угроз жизни человека в связи с легализацией прав роботов на убийство существует ещё одна проблема, которую необходимо осветить в этой работе. Этот вопрос связан с тем, как меняются условия ведения военных действий. Они меняются и для стран, использующих ударные БПЛА и другие формы боевых роботов, и для тех, на кого направлены эти силы.

Во-первых, мир сталкивается с неравенством в вооружении боевыми роботами среди возможных участников боевых действий. Существуют страны-производители вооружения, например США и Израиль, Южная Корея. Существуют страны, которые выделяют бюджетные средства на закупку роботов или на развитие военной робототехники, Россия, Индия, и те, кто не обеспечивает себя таким вооружением. Как можно оценивать нападение на страну, не имеющую боевых роботов, технически более оснащённой страной-агрессором с точки зрения справедливого ведения войны? Международное гуманитарное право пока не включило роботов в число объект регуляции при ведении военных действий, а в Женевской конвенции по защите жертв войны роботы не числятся, как запрещённое оружие. Тем самым, правовые документы, проводящие границы и этически дозволенного минимума в правилах ведения войны на данный момент не регулируют процесс. Первым шагом к решению этой проблемы могут стать изменения в международных документах, регулирующих гуманитарное и военное право.

Вопрос коммерциализации войны и робота, выступающего в роли наёмника, так же привлекают обсуждения и споры. Компания «Samsung» из Южной Кореи и канадская частная компания «Clearpath Robotics»

занимаются и военными роботами, и робототехникой для военных решений. Их продукты может купить другая страна и использовать для обороны своих границ или ведения активных боевых действий, так робот начинает помимо роли продукта, товара восприниматься, как наёмник.

Вопрос статуса робота переносится на статус всей армии, она становится наёмной, ведь на её службе роботы, купленные за бюджетные деньги, словно солдаты-контрактники, которые служат за определённую сумму. В чём состоит несправедливость в существовании наёмной армии как таковой. Первый аргумент против наёмной армии заключается в том, что страдает связь государства и гражданина. Служение в армии – это гражданская обязанность, упоминание о ней есть в главных законодательных актах страны, например в Конституции РФ. Перекалывание обязанности гражданина на наёмную силу ведёт к отказу от гражданского долга.

М. Сэндел считает, что наёмники превращают военную обязанность не в долг, а в товар, что ведёт разложению гражданских идеалов¹⁹. Это нивелирует возможность гражданина страны реализовать гражданскую добродетель – защищать свою страну. Так разрушаются границы морального облика гражданина. Это вторая особенность проблемы. Так же покупкой ударных дронов или боевых роботов могут заниматься не только страны, но и террористические группировки, что несёт глобальную проблему безопасности. В этой ситуации ориентироваться можно на принципы робототехники, которые выступают за производство военных роботов только государством и за государственную ответственность за ведение боевых действий с помощью роботов. Возможно, распространение и узаконивание этих принципов станет первым шагом на пути к решению этого вопроса.

Проблема компьютеризации войны – это ещё один аспект, который значительным образом меняет реальность боевых действий. В масштабах государства с технологиями военные действия становятся более безопасными, количество человеческих потерь значительно снижается, следовательно

¹⁹ Сэндел М. Что такое справедливость «Манн, Иванов и Фербер»: Москва; 2013. 108 с.

вопрос ведения войны может остаться за наличием вооружения и денег на покупку роботов и поддержания их состояния.

При этом можно взглянуть на компьютеризацию со стороны тех, кто осуществляет их. Решения о нанесении ударов, убийствах врагов принимают операторы. Операторы опосредованным образом находятся на поле боя с помощью техники, но нельзя сказать, что они не погружены в реальность сражения, операторы обладают таким техническим снаряжением, мониторами, микрофонами, связью и т.д., что на обвинения в том, что такая война похожа на видеоигру, критики получают такой ответ от операторов БПЛА : «Мировосприятие Nintendo — это совершенно отдельное мировосприятие, «как в видеоигре», которое является «обманчиво стерильным». — Эта штука реальна. Я по-настоящему лишаю людей жизни. Я стреляю из реального оружия. И я должен нести ответственность за свои действия в реальности»²⁰. БПЛА в отношении военных становится и новым орудием, и методом убийств на расстоянии, они воздействуют на общественную мораль и этические принципы оператора. Помимо вопроса компьютеризации методов ведения войны, операторы столкнулись с тем, что их деятельность оценивается обществом и ими самими по-разному. Общество не одобряет и нивелирует деятельность оператора, которого называет «диванным солдатом» и приписывает ему состояния офисного работника. Однако сами операторы, помимо того, что они осознают ответственность за свои действия, переживают серьёзные стрессы, посттравматические синдромы сравнимые с теми, что получают военные, находящиеся в реальных боевых действиях. В этом вопросе можно ориентироваться на опыт операторов БПЛА, которые понимают опасность и угрозы боевых действий и остаются чувствительными к последствиям нанесённых ими действий, а не на государство, которое опирается на статистику количества поражённых противников при сохранности живых бойцов, защищающих страну.

²⁰ Минделл Д. Восстание машин отменяется! Мифы о роботизации: «Альпина нон-фикшн»; 2017. 102 с.

В качестве вывода к параграфу, посвящённому боевым роботам можно сказать, что перед лицом государств и граждан стоят реальные проблемы статуса робота в качестве товара, наёмника, дееспособного лица. Так же вопросы сохранения чувствительности к ценности жизни среди стран и военных, имеющих возможность поражать противника на расстоянии, чтобы не превратить войну в игру. Учитывая наличие локальных конфликтов по миру и активное развитие боевой робототехники, необходима работа с декларативными и международными документами о способах и методах справедливого использования боевой робототехники и определении субъектов, имеющих право на её производство.

Понятие социального робота

Роботами-помощниками (Service Robots) принято называть роботов, которые действуют полу- или полностью автономно, выполняя при этом работу, направленную на улучшение благополучия человека, помимо этого их действия направлены на улучшение состояния оснащения и аппаратуры учреждения. Однако процесс поддержки какого-то оборудования роботами-помощниками исключает выполнение какого-либо обслуживания на промышленных производствах.²¹ Это отличает их от роботов и роботизированных систем, занятых в промышленном секторе.

Роботы-помощники делятся Международной организацией Робототехники на несколько видов. Первое деление связано с разделением сфер использования таких машин на частную (домашнюю) и профессиональную. Второй фактор, разделяющий роботов – это сфера деятельности.

В сфере домашнего использования роботы выполняют домашние поручения (робот-пылесос для поддержания чистоты или робот-компаньон для налаживания социальных аспектов жизнедеятельности), являются средствами развлечения (роботы-игрушки), помогают престарелым и

²¹International Federation of Robotics URL: <https://ifr.org/service-robots/> (дата обращения: 08.03.2018)

инвалидам (экзоскелеты, инвалидное кресло), занимаются транспортом маломобильных людей, охраняют жилище.

В профессиональном использовании разнообразие роботов велико, можно лишь перечислить отдельные области: сфера обслуживания и туризма, здравоохранение, транспорт, сельское хозяйство и животноводство, охрана и логистика.

Роботы в медицине

Использование роботов непосредственно в медицинских операциях уже не является новинкой, например на протяжении 10 лет в хирургии используют полуавтономных роботов для борьбы с раком простаты. Положительными сторонами проникновения роботов в эту сферу человека можно назвать отсутствие потери в эффективности операции, проведённой роботом или хирургом. Отрицательные элементы касаются проведения инвазивных действий с помощью робота, который может отображать лишь визуальную часть вмешательства, когда человек-хирург лишён тактильного контакта, который не может обеспечить робот. Помимо этого роботы являются дорогостоящими и их обслуживание так же ресурсозатратно, что во многом тормозит повсеместное внедрение технологий в больницы. Среди этических проблем в медицине можно выделить проблему распределения роботов в системе общественного здравоохранения, количество обычных хирургов и вопросы ответственности за хирургические практики без роботов и с ними.

Помимо такой классификации учёные предлагают разделить роботов согласно их функциональным особенностям. Выделяют три типа робототехники: вспомогательная, социально интерактивная²² и социально ассистивная²³. Первая направлена на помощь людям с помощью физических взаимодействий, вторая использует социальные качества, как ключевой

²²Fong T. Nourbakhsh, I. & Dautenhahn, K. A survey of socially interactive robots.//Robotics and Autonomous Systems (2003) Vol.42, pp.144.

²³Feil-Seifer D. Mataric M. J. Defining Socially Assistive Robotics. Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics. (2005) USA, pp.467.

элемент деятельности, где выступает в качестве партнёров, помощников и сверстников для пользователя. Третий тип робототехники объединяет первые два между собой и синтезирует их способности помощи человеку, ориентируясь на возможности социальных взаимодействий.

Несмотря на развитие сферы развлечений, технологий умного города и умного дома, которые функционируют с помощью роботизированных систем и упрощают человеческие заботы, кажется более актуальным, остановиться на роботах-помощниках, которые больше всего влияют на благополучие человека. Такие роботы заняты и в частной, и в профессиональной сфере и направлены на поддержание здоровья человека.

Главным фокусом в этом делении становится степень взаимодействия между роботом и человеком. Взаимодействие – это понятие, отражающая процессы воздействия различных объектов друг на друга, их взаимную обусловленность, это вид отношения или связи²⁴. Раньше существовал лишь термин, отражающий способ выстраивания отношений между человеком и компьютером «human-computer interaction». Существовал определённый цикл взаимодействий между человеком и компьютерной техникой. В этом цикле человек задаёт характер отношений путём определённого действия – набора данных, регуляции кнопок или рычагов на устройстве, всё это делается ради достижения цели. Для того чтобы установилось взаимодействие устройство действует и в конце подаёт сигнал в виде звука, включение света и т.д. – отзывается и сообщает о текущем состоянии. Перед тем как ответить, человек анализирует полученные данные и сравнивает то, что получилось и то, на какой результат было направлено действие изначально. Так циклы могут повторяться множество раз. Такие отношения строятся на законах робототехники, и социум никак не может повлиять на них. Однако в человеческом обществе общение не происходит с помощью физики и

²⁴Ильичёв Л. Ф., Федосеев П. Н., Ковалёв С. М., Панов В. Г. Философский энциклопедический словарь. — М.: Советская энциклопедия. 1983. С. 365

математики, оно имеет социальную природу. Выделяются несколько признаков социального взаимодействия

1. Всегда предполагает обмен символами, знаками, которые принимающий может раскодировать. Для коммуникации в этом случае подходят слова в устной речи, эмоции, письменный язык, язык жестов, который понятен обеим сторонам при взаимодействии.

2. Социальное взаимодействие имеет цель, обусловленную намерениями участников. Здесь нужно отметить, что деятельность является не запрограммированной или инстинктивной, она сознательная и несёт элементы целеполагания, выполняется для удовлетворения какой-то потребности.

3. Это внешняя связь участников, доступная для наблюдения. Совместное занятие деятельностью, беседа – активно задействуют человека и предполагают внешнее проявление, а не только работу сознания.

4. Такое взаимодействие связано с конкретной ситуацией. Поддержание беседы, способность заинтересовать каким-то делом, помочь справиться с заданием, обусловлено средой, целью и конкретными субъектами.

Так постепенно среди «human-computer interaction» выделилось направление «human-robot interaction», направленное на изучение того как могут взаимодействовать робот и человек. При этом изначально считалось, что получение данных о человеке (пол, возраст, настроение, социальный статус т.д.) являлись препятствием для функционирования робота. Область исследования взаимодействий робота и человека стала не просто ступенью развития знания, а необходимостью, потому что появились социальные роботы. (Social Service Robots).

Социальные роботы начали разрабатываться в начале XX века. Прежде всего, их разработка была направлена на привлечение публики, развлечения. Способности таких роботов были примитивными на сегодняшний день:

движение головой и рукой, сидение, стояние и даже выполнение простых команд. Так разработчики присваивали роботам плохие привычки, например курение сигар или забавные, надувание воздушных шаров. В истории робототехники существуют интересные примеры антропоморфных социальных роботов. Робот Мистер Телевокс, один из первых антропоморфных роботов, умел изящно выкуривать сигары и был примером джентльмена, а робот Электро словно человек стал обладателем питомца, робопса Спарко.

В 40-е года XX века стала популярной идея искусственного интеллекта из-за выхода в свет книги «Кибернетика, или управление и связь в животном мире» математика Норберта Винера. Для разработчиков роботов поворотной мыслью, изложенной в книге, стало то, что роботу необходимо выражать обратную связь и взаимодействовать с окружающей средой.

Здравоохранение и роботы ассистенты

Объектом заботы робота-помощника может стать любой человек, однако в Международном отчёте ЮНЕСКО по робоэтике выделяют две основные группы, для которых терапия с участием роботов развивает более активно.²⁵

Первая группа – это дети с задержками в развитии. Роботы такого типа занимаются с детьми с синдромом Дауна, работая над улучшением их обучаемости. Роботы «Nao» человекоподобный робот, «Pleo» в виде динозавра и «Tito» в виде антропоморфной игрушки участвуют в эксперименте, в котором он занят развитием адаптивности у детей с аутизмом уже в 2012 году с помощью техник социального взаимодействия захвата и поддержания внимания ребёнка, подражания его действиям²⁶.

Аутизм – это комплексное заболевание, прямые причины которого нельзя выявить и вылечить. Стереотипное поведение, ограниченность

²⁵ Report Of Comest On Robotics Ethics//UNESCO; 2017 pp.35.

URL :<http://unesdoc.unesco.org/images/0025/002539/253952E.pdf> (дата обращения: 10.03.18)

²⁶ Scassellati B. Admoni, H. and Mataric M. Robots for autism research// Annual Review of Biomedical Engineering, No.14 (2012) pp. 279-280.

интересов и повторяемость движений – распространённые симптомы этой болезни. Такие дети имеют проблемы с коммуникациями, т.к. монотонность речи и неуместные обороты речи затрудняют налаживание общения. Для улучшения социальных и языковых способностей таких детей над ними работают команды учёных, представляющих различные отрасли науки. Так же представлены исследования, где на базе эксперимента с такими детьми занимался робот, направленный на улучшение приспособляемости детей к окружающему миру. Дети-аутисты чаще всего были увлечены роботами из-за того, что влияние робота на ребёнка стабильно, а движения медленные и предсказуемые, в отличие от поведения и эмоций человека.²⁷

Такие занятия роботов с детьми на данный момент нельзя сделать стандартной процедурой и внедрить во все больницы и детские центры. Этому мешает разнообразие проявления симптомов у детей с аутизмом и синдромом Дауна, т.е. нельзя подобрать универсальный алгоритм взаимодействия с ребёнком. Так же врачи имеют сомнения в утверждении, что взаимодействие ребёнка с роботом можно назвать общением. В такой коммуникации в слабом виде проявляется эмоциональная (отдача робота) сторона общения, без неё ребёнок не сможет учиться реагировать на эмоции других людей. Тем самым, ученые и врачи стоят на позиции, что такой вид терапии является промежуточным и не может быть окончательным.

Демографические изменения, которые касаются общества в развитых странах мира, позволяют прогнозировать большой рост использования технических решений для проблем, связанных с уходом за пожилыми людьми. Исследования в области робототехники, связанные с заботой и уходом за старшим поколением, являются одними из самых многообещающих. Этому способствует общая тенденция роста количества пожилых людей и старения населения в европейских странах, Японии и США. Социологические исследования 2010 года прогнозируют, что к 2060

²⁷Scassellati B. Admoni, H. and Mataric M. Robots for autism research// Annual Review of Biomedical Engineering, No.14 (2012) pp. 280.

году процент людей старше 65 лет в странах Европейского союза вырастет с 17,4% до 30,0%.²⁸ Исследования Евростата в июне 2017 года показывают рост пожилых людей старше 65 лет до 19,2%.²⁹ Прогнозы социологов медленным темпом находят своё подтверждение в реальной жизни.

Обслуживание человека техникой, например кофейным автоматом или стиральной машиной, вызывает не так много вопросов, связанных с этикой. Робот же и его использование непосредственно связаны с человеком и выявляют много спорных или проблемных позиций, касающихся характера этих взаимодействий.

Для более предметного описания актуальных проблем машинной этики в области ухода за пожилыми людьми было решено взять в пример робота Генри. Он часть европейского научного проекта STRANDS³⁰. Генри автономен, мобилен, среди его внешних качеств можно выделить антропоморфные черты и зелёный цвет его покрытия. Среди задач и возможностей робота – информирование через сенсорный экран пациентов о том, что их вызвали на приём, направление посетителей в нужные отделения и офисы (Австрийский институт здравоохранения). Генри занимается патрулированием коридоров. Он сопровождает прогулочные группы во время физиотерапии, играет музыку или издаёт звуки и показывает изображение на своём экране, чтобы стимулировать включенность людей, страдающих деменцией. Отмечается, что до этого никто из персонала института здравоохранения не видел и не взаимодействовал с социальным роботом до этого момента.

На примере Генри венский учёный, Т.Кörtner, сделал обзор очевидных вопросов и этических проблем, связанных с использованием Генри.³¹ В

²⁸Demography report 2010. Older, more numerous and diverse Europeans. European Commission, March 2011. URL: http://epp.eurostat.ec.europa.eu/cache/ITY_OF_FPUB/KE-ET-10-001/EN/KE-ET-10-001- EN.PDF (дата обращения: 09.03.2018)

²⁹Eurostat: Population structure and ageing URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing (дата обращения: 09.03.2018)

³⁰STRANDS project URL:<http://strands.acin.tuwien.ac.at/> (дата обращения: 29.03.18)

³¹Körtner T. Ethical challenges in the use of social service robots for elderly people//Zeitschrift für Gerontologie und Geriatrie No.4 (2016) pp. 304.

области развития роботов социальных или роботов-компаньонов большой упор делается на копирование внешности и поведения человека, как и Генри, они ориентируются в окружающей среде и достаточно автономны, чтобы выводить независимые решения. Помимо этого роботы становятся связанными с Интернетом, они проводят различные манипуляции с потенциально личными данными людей.

Кёртнер, наблюдавший за Генри, опирался на некоторые установки, касающиеся моральной философии.

Во-первых, он разделяет понятия морали и этики, для него мораль – это механизм, который производит оценку, используя добро и зло как полярные оценки на шкале поведения субъекта деятельности.

Этика – это теория морали, а так же в большей своей сути представляет этос. Этос для него индивидуален и сильно завязан на личности человека и его отношении ко всему. Он так же выделяет изменчивость как качество этики и считает, что её функция – разработка лучших теорий для блага всего человечества.

В связи с этим возникают вопросы и разные точки зрения на то, какая основа должна лежать в машинной этике. Т.Кёртнер поддерживает теорию, что машинная этика должна исходить из ценностей и принципов, которые провозглашает международное право.

Эту позицию продвигает исследователь Ф. Оперто, он предлагает идею, где машинная этика и робоэтика имеет соответствие с международными документами: Всеобщей декларацией прав человека и Лиссабонским договором, потому что они составляют основу для декларирования ценности человеческого достоинства³². Уважение личности и достоинства, утверждение всеобщих прав человека – это база, на которую необходимо опираться, формируя машинную этику и её принципы.

Среди исследователей, которые пытаются сформулировать список универсальных этических принципов, можно увидеть тенденцию

³²Operto F. Ethics in advanced robotics.//IEEE Robot Autom Mag 1 Vol.18 (2011) pp.72–78.

ориентироваться на принципы биомедицинской этики, так исследователи Китченер и Андерсон выделили 4 основных принципа этики³³:

1. Не навреди
2. Благодеяния
3. Уважения к автономии личности
4. Справедливости в распределении ресурсов

В сфере биомедицинской этики ещё выделяются принцип честности, справедливости и информированного согласия. Кажется, что социальным роботам-помощникам можно вменить следование этим принципам (запрограммировать их таким образом), тем самым удовлетворить притязания со стороны медицинских учреждений и притязания к вопросу об универсальности этих принципов. Однако стоит обратить внимание на Генри, социального робота-помощника, и на практике выяснить, почему решение сделать кальку принципов, сформулированных для людей, и применить её к роботам не является полностью верным решением.

При взаимодействии с роботом возникают ситуации, где строго говоря, такие принципы встретить нельзя. Как научная платформа Генри не тестировался для всех ситуаций и подвержен ошибкам. Например, принцип благодеяния: при том, что принцип является целью, он не всегда может быть достигнут.

Примером этому служит то, что Генри со своим сенсорным экраном подходит для людей, которые крепко стоят на ногах, но не всегда доступен для тех, кто использует инвалидное кресло. Здесь справедливое распределение блага обеспечивается не для всех и не в любой ситуации. Это показывает, что, несмотря на следования этическим принципам, можно быстро столкнуться с проблемой этического характера. Тем не менее, только при тестировании технологий в реальном мире можно добиться прогресса в этой сфере. Для того, чтобы узнать считают ли пользователи и управляющие

³³Kitchener K.S. Anderson S.K. Foundations of ethical practice, research, and teaching in psychology and counseling (2011) Routledge, London p. 144.

роботом его полезным, прототипы должны быть протестированы в уязвимых группах. Термин «уязвимые группы» используется в отношении детей, отдельных категорий женщин, инвалидов, трудящихся-мигрантов, беженцев, национальных меньшинств и коренных народов, пожилых людей. В этой части исследования нас интересуют только пожилые люди, как уязвимая группа.

Ожидается, что роботы, заботящиеся о пожилых людях, будут приносить пользу для всех участвующих, но риск всегда предполагается. На опыте Генри по взаимодействию с пожилыми пациентами, дееспособными и недееспособными, разработчики выделили несколько актуальных проблем и ситуаций, вызывающих риски, которые предлагается проанализировать.

В литературе и опыте проекта по пространственно-временным представлениям и деятельности для когнитивного контроля в долгосрочных сценариях (STRANDS), текущими этическими проблемами можно назвать такие:

- Введение в заблуждение
- Достоинства
- Изоляции (оторванности\изолированности)
- Защиты данных
- Конфиденциальности
- Безопасности

Введение в заблуждение:

Прежде всего, стоит сказать, что этот термин предлагается понимать как самообман в результате восприятия робота, как личности и выстраивание доверия и социальных ожиданий, не как от машины, а как от человека. Рост и дизайн, внешние черты робота могут вызывать у человека определённые ожидания относительно функциональности и уровня интеллекта этого робота. Помимо этого, на наши представления очень влияет речь, она является ключевым фактором, т.к. с помощью неё можно осуществлять коммуникативное взаимодействие между машиной и человеком. Роботы

имеют возможность производить запрограммированную речь и распознавать человеческую. В зависимости от того, как их задумали разработчики и инженеры, робот может производить звуки, так он будет восприниматься людьми, как машина. Однако они так же могут производить вежливую речь, направленную на собеседника и создающую диалоговую форму «общения», при таком обмене информацией у робота в лице человека возникает особая личность.

Генри, как было замечено, только расположившись на территории, которую он патрулировал столкнулся с тем, что люди ожидали от него человекоподобного общения, диалога и понимания человеческого языка. Генри так не умеет, и узнав это, люди очень расстраивались, что нельзя наладить контакт с роботом.

Другие роботы создаются с дизайном, способным передавать эмоциональные реакции. В связи с прогрессом сложности и автономности робототехники, есть опасения о постепенном размытии грани между машиной и реальным человеком. При этом уязвимые группы людей, в том числе дети и пожилые люди, могут быть подвержены этому размыванию границ больше остальных членов общества. Всё из-за того, что они имеют склонность проецировать эмоциональные способности на машину, с которой они взаимодействуют, когда этой эмоциональности нет. Тёркл заметил, что пожилые участники исследования эмоционально привязывались к роботу³⁴. Это вызвало неправильные предположения о интеллектуальных и эмоциональных возможностях робота. Некоторым было свойственна тревога о том, что робот будет скучать, пока хозяев нет дома. Это приводит к вопросу: являются ли роботы средством для введения в заблуждение?

Обман, введение в заблуждение человека машиной рассматривается участниками проекта, как неэтичное поведение, потому что разработчики и сам робот не могут быть полностью ответственны за проблемы восприятия,

³⁴Turkle S. Relational artifacts with children and elders: The complexities of cybercompanions//Connection Science18 Vol.4 (2006) pp. 359-361.

но и пользователя робота нельзя обвинить в том, что он не правильно воспринимает робота. На данный момент, все роботы не имеют возможности обладать эмоциональностью и сознанием. Однако человеческая натура склонна наделять человеческими качествами объекты от любимых игрушек, до машин или компьютеров, это стремление наполнить нечеловеческих агентов (актеров) характеристиками личности, мотивациями, намерениями, эмоциями. Тем больше становится вопрос о дизайне и производстве антропоморфных роботов для ухода за пожилыми людьми, при том, что есть вероятность получения стрессов и обретения ложного доверия к машине.

«Выводы и этические рекомендации»: для того, чтобы решить проблему самообмана необходимо информирование о возможностях и пределах активности роботов пользователей машин, чтобы избежать искажённого восприятия машины. Проведение информативных бесед среди пользователей с ответами на вопросами о правилах взаимодействия с роботами. Сотрудничество с медицинским персоналом не только пациентов, но разработчиков и психологов для избегания будущих последствий в виде стрессов и тревоги пациентов при создании роботов и алгоритмов для их работы.

Достоинство

По части заботы и терапии среди пожилых людей самый именитый на сегодняшний день – это робот-тюленёнок PARO. Это машина, повторяющая образ живого тюленёнка: мягкий покров из белой шерсти, маленький вес, чтобы было возможно взять его на руки. Он издаёт звуки малыша тюленя, у него моргающие большие чёрные глаза, он тёплый, создаёт вибрации и двигает телом подобно процессу дыхания и реакции на обращение к нему: поворачивается, поднимает голову с туловищем, двигает глазами и плавниками. Его задачей является вовлечь пожилых пациентов с деменцией во взаимодействие и таким образом включить их в активность и мотивировать к ней, дать им позаботиться о питомце. Среди всех нововведений, улучшений и эффектов не просматриваются проблемы,

которые проявляются в инфантилизации пожилых, страдающих деменцией и снисходительного обращения с ними. Инфантильность – это незрелость или неразвитость психики.

Инфантильный человек – это человек, в поведении которого доминирует незрелое поведение, неготовность брать ответственность на себя и самостоятельно принимать решения, отсутствие жизненных целей и стремление что-то поменять в себе и в своей жизни в целом.

Инфантильное расстройство личности подразумевает под собой наличие у взрослого человека черт и поведения, характерных для ребёнка. «Инфантилизация» среди уязвимых групп из-за нарушения этим прав человека считается неэтичной в гериатрии. Способствование проявлению беззащитности и нежелания брать ответственность за себя во многом умаляет достоинство человека и способствует развитию мировоззрения, где человек не способен понимать ценность своей собственной жизни.

Это можно проследить на примере Генри: когда пациенты группы следуют за ним по коридорам и поют традиционные песни, которые он играет, они лишь инфантилизируются или всё же получают эмоциональное удовлетворение от этой деятельности? В настоящее время вопрос о роботах, используемых в терапевтических целях, остаётся открытым.

«Выводы и этические рекомендации»: контроль этической комиссией взаимодействий роботов и пожилых (больных деменцией, недееспособных и т.д.). Привлечение медперсонала, родственников на раннем этапе взаимодействий с роботами, чтобы проинформировать, успокоить и наглядно показать как происходит взаимодействие. Информационные мероприятия, семинары, обратная связь, контакты с теми, кто пострадал от взаимодействия, все эти шаги будут влиять на решение проблемы. Это способствует увеличению доверия к робототехнике и увеличению лояльности к использованию робототехники среди пациентов.

Изоляция:

Изоляцию называют одной из главных проблем, которая встаёт на пути использования роботов, особенно в уходе за пожилыми. Социальная забота за уязвлёнными людьми должна остаться в руках машин, по мере их усложнения и развития. И здесь есть зависимость: чем больше мы взаимодействуем с роботами, тем меньше мы имеем возможности общаться с другими людьми, особенно это актуально для пожилых людей, которые могут быть ограничены в своей мобильности. Когда существуют доказательства того, что социальные связи и дружеское общение помогают задерживать начало симптомов деменции, существует нехватка доказательств и опыта долгого взаимодействия с роботами, показывающих такие же результаты.

Кажется, что лучше признать опасной нехватку межличностного общения при уходе за пожилыми с помощью роботов. Неэтичными нововведения можно назвать в таком случае, когда более значимой целью становится увеличение независимости роботов и улучшение их способностей, чем сам уход за людьми. Именно этот перевес считается выходящим за пределы нормы.

При этом споры разработчиков и инженеров, выступающих против этой точки зрения, не утихают. Они спорят о возможностях технологических устройств и роботов, способствующих улучшению коммуникации между людьми.

В разрезе этой проблемы, хотелось бы добавить, что она тесно связана с психологическими проблемами пожилых пациентов. Стоит подчеркнуть тенденцию исследователей, анализирующих поведение социальных роботов, опираться на эмоциональный аспект взаимодействий, смещая этические принципы в область проблем психологического взаимодействия робота и человека. Проблема изоляции так же может скрываться в эгоистическом направлении деятельности системы здравоохранения. Система рассматривает человека, как единицу, отправляя алгоритмы работы с пациентами в больницы. Система направлена на достижение средних результатов,

опирается на статистические данные. Оптимизация процесса при которой почти все обязанности по уходу за пожилыми будут выполняться роботами-помощниками, освобождая мед.персонал для более сложных задач, могут оставить пожилых без альтернативы человеческого или технического взаимодействия, особенно тех, кто лишен возможности передвигаться или не имеет родственников, опекунов или друзей.

Принцип благодеяния и ценность заботы о человеке останется только «в руках» робота-помощника, который не сможет транслировать эту ценность, тем самым оставляя её только в теории.

Конфиденциальность

Сейчас при существовании телевидения и связи людей через Интернет несмотря на расстояние, границы открытой и частной информации становятся более гибкими, размытыми и их труднее определить\различить. Для того чтобы получить выгоды от лечения врачи разглашают личную информацию для информационной базы и роботы, ухаживающие за пожилыми не являются исключениями. Генри оснащён различными камерами, которые могут хранить фото и видео материалы. Они нужны ему для того, чтобы нормально двигаться в пространстве, различать людей и препятствия. Эти видео данные и изображения проходят обработку, а так же могут быть сохранены у других роботов. Кроме этого существуют датчики, реагирующие на движение и измерение активности. В этой ситуации пользователи вряд ли понимают, что происходит с этими данными.

Выявленная зависимость, что человек, более нуждающийся в заботе, с большим желанием поступаются своей конфиденциальностью. Несмотря на то, что запись жизненных показателей или запись в туалетных комнатах с помощью камер были наименее приемлемы пациентами, всё равно сенсорные технологии приемлемы для большинства людей. Такие результаты предполагают то, что пожилые люди будут вынуждены принять роботов и привыкнуть к роботам, которые передают их личную информацию. Из этого следует, что пожилым пациентам придётся понять основы работы

технологии, чтобы сохранить свои права на автономию. Система сбора данных и хранения должна быть прозрачной. Необходимо сделать понятным какая информация собирается, когда у кого есть к ней доступ.

В ситуации учреждения и установок, где находится Генри, нужно учитывать не только интересы пациентов, но и работников, посетителей, которые не могут не замечать, что являются объектами наблюдения робота. Даже в этом случае конфиденциальность и защита данных не могут быть гарантированы. Т. Дейннинг обнаружил, что простейшие роботы, выполняющие работу по дому, могут быть легко обнаружены и идентифицированы удалёнными пользователями, так же они могут быть взломаны ими, что даёт возможность управлять ими и контролировать³⁵. Чем совершеннее робот, тем задача в сфере интернет безопасности становится сложнее. Потеря контроля и автономии может вызывать чувства объективации у пользователей, что усиливает важность человеческого контроля над роботом. При работе в общественной, государственной или полугосударственной сфере шагом к решению проблемы может стать регистрация в агентстве по защите данных, так же как и было сделано с Генри.

«Выводы и этические рекомендации»: отключение функций записи видео там, где обозначена личная сфера пациента. (не снимать там, где проходят процедуры или их ожидание) В домашнем использовании нужны настройки и возможность определять зоны, которые не должны входить в запись.

Информированное согласие в мед.учреждении – сложно и почти невыполнимо взять его у всех, однако следует информировать о наличии робота, записи данных и целях, в соответствии с которыми проводится запись. Подчинение местному законодательству и международным декларациям по правам человека – это шаг к решению проблемы со стороны

³⁵Denning T. Matuszek C. Koscher K. Smith JR. Kohno T. A spotlight on security and privacy risks with future household robots: attacks and lessons.//In: Proceedings of the 11th international conference on Ubiquitous computing 11th international conference on Ubiquitous computing. (2009) pp.105–114.

правового регулирования. Регистрация в агентстве по защите данных, получение разрешения на сбор данных в больнице поможет сохранить конфиденциальность данных и обеспечить гарантии для тех, кто может стать потенциальной жертвой киберпреступлений или утечки данных.

Для защиты данных – нужна юридическая регистрация, одобрение этическим комитетом. Полная защита всех действий с данными: передачи, переработки, хранения должны регламентироваться. Жёсткие диски должны быть оснащены сложным паролем, не подключены к интернету и быть доступны ограниченному кругу лиц – всё это возможно отразить в кодексах поведения и корпоративных документах медицинского учреждения.

Безопасность:

Роботы могут наносить ущерб, однако только в случае военных роботов – это можно трактовать как запланированный результат активности робота. При разработке мирных роботов цель нанесения вреда отсутствует полностью, чтобы обеспечить необходимую безопасность окружающих. Зато на практике прототипы роботов показывают, что при всех усердиях инженеров, все случаи возникновения опасности вследствие активности робота нельзя избежать. Генри, курирующий обстановку в коридоре, завернув за угол, столкнулся с человеком, который помогал человеку в инвалидном кресле. В другой раз Генри упал с лестницы. Один из актуальнейших вопросов: на ком лежит ответственность за вред, нанесённый роботом? Ответственным лицом может быть как разработчик, дизайнер, так и компания, в которой он был создан. Стоит ли заводить дискуссию о личной ответственности робота за свою активность. Все эти проблемы предлагается решить с помощью работы страховых компаний, т.к. на сегодняшний день дискуссии о личной ответственности робота – это только дискуссии, а необходимость обеспечить гарантиями пользователей роботов или их прототипов существует на данный момент. Так же разработанные международные принципы по робоэтике в 2010 году говорят о том, что человек должен убедиться, что активность робота соответствует всем

законам страны и принципам международного права, перед тем как вводить робота в эксплуатацию³⁶. Тем самым они декларируют отсутствие у робота личной ответственности. Следовательно, можно сказать, что сейчас главным образом стоит вопрос, как организовать работу страховых компаний, в соответствии с этическими принципами и идеалами безопасности. (в мед.учреждении) При этом нужно отвести особое внимание к тому, что у обычного человека и представителя уязвимых групп может складываться разное ощущение безопасности при взаимодействии с роботом.

«Выводы и этические рекомендации»: проведение пилотных тестов – обязательно, необходимо и чрезвычайно важно для обеспечения безопасности персонала, пользователей робототехники. Определение запретных зон для робота, чтобы он не провоцировал опасные ситуации, а персонал знал, где находятся зоны работы машины. Предусмотрение внештатных ситуаций и обучение сотрудников взаимодействовать с машиной при неполадках и сбоях. От разработчиков зависит предсказуемый характер поведения роботов для персонала и клиентов, что так же стало решением вопроса безопасности. Обратная связь от пациентов, клиентов и работников для улучшения предсказуемости поведения и избегания аварий и недоразумений так же станут катализаторами для позитивных результатов в вопросах безопасности.

Уязвимость:

Нельзя не учитывать уязвимость пожилых людей в качестве участников робототехнических исследований и пользователей роботов-помощников. До тех пор, пока участники могут понимать цель технологии и могут выражать свои намерения, им предлагается давать информированное согласие на проведение каких-либо исследований. Если пожилой человек не является дееспособным, его болезнь или состояние не позволяет ему отдавать отчёт в своих действиях, то тогда регулировать участие в исследованиях

³⁶ Principles of robotics//EPSRC URL:

<https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/> (дата обращения: 14.03.18)

предлагается с помощью информированного согласия, данного доверенным лицом, опекуном. Скорее всего, эта процедура будет занимать долгое время при обработке, но только при её наличии и заверении можно будет приступать к взаимодействию с робототехникой.

При отсутствии чёткого регулирования в виде информированного согласия, нужно чётко объяснить участнику исследования обоснованные риски по сравнению с выгодами, которые могут быть получены в ходе взаимодействия с робототехникой. Всё это необходимо, т.к. влечёт стрессы и напряжение для участников. Взаимодействие с роботами может акцентировать внимание на слабостях и физических ограничениях, которые испытывает пациент, что может сказаться на уровне агрессии, неудовлетворённости и изнурительном характере прохождения испытаний. Этические характеристики поведения социального робота-помощника всегда должны опираться на человеческий фактор, потому что появление робота в жизни человека влияет на его (душевное, ментальное, эмоциональное) состояние.

«Выводы и этические рекомендации»: Социальные роботы-помощники должны быть сфокусированы на возможностях и ограничениях уязвимых групп. Не каждое техническое решение подходит каждому и важно обращать внимание не только на тех. совершенствование робототехники, но и на снижение порога тревожности и стресса при использовании робота. Нужно уважать сознательный выбор человека не использовать робота или его определённые функции.

В целом можно сделать вывод, что социальная ассистивная робототехника активно внедряется в процессы ухода и взаимодействия с разными пациентами. Особенности больных влияют на дизайн, функции, строение и оснащение роботов. Нынешние образцы такой робототехники вызывают проблемы безопасности, конфиденциальности данных, затрагивают вопросы уязвимости и самообмана отдельных категорий пациентов, заставляют учёных создавать технические решения, которые не

будут способствовать изоляции пациентов и при этом заботиться о достоинстве и уважении каждого участника взаимодействий с роботом.

Среди военной и социальной робототехники большинство проблем не перекликались, а остались уникальными для своей области. Однако так или иначе, этические проблемы обеих сфер касались ценности человеческой жизни, уважения и достоинства пользователя робототехники, а так же вопросов безопасности. Конечно, стоит отметить тенденцию не только к осмыслению этических проблем применения робототехники, которая активизируется с появлением новшеств в индустрии, но и медленного движения в сторону решения этих проблем в виде составления проектов кодексов, принципов, декларативных документов, формирования обществ выступающих за или против боевых роботов. Всё это говорит о том, что общество озабочено проблемами машинной этики и готово принимать участие в решении актуальных проблем.

Заключение

Целью работы было представление текущего этапа развития одного из направлений этики ИИ, для этого были выбраны преимущественно современные источники англоязычных авторов, которые максимально отразили текущее состояние робототехники и этической мысли. Было выяснено, что машинная этика – это область моральной философии, которая занимается разработкой подходов к установлению механизма принятия этических решений интеллектуальными машинами и системами. Одно из важных свойств направления выделенных в работе, что машинная этика имеет междисциплинарный характер и ориентируется на разработки робототехники и других компьютерных наук.

Помимо этого в работе были сформулированы понятия ИИ, работа. Под роботом следует понимать механизм, с определённым набором компонентов. Удалось установить то, что ИИ – это одновременно и научная сфера, и способность.

Так же было необходимо представить современное понимание проблем теоретического и прикладного характера, которые проявляются в обществе с приходом новых технологий и формируют облик машинной этики. В теории – это проблемы морального действия агентов, проблема искусственной морали. Эти проблемы были проанализированы с помощью применения на них этических теорий.

В ходе работы установили, какие проблемы возникают у людей, которые непосредственно имеют дело с робототехникой. Это было сделано на примере социальной и боевой робототехники, выяснилось, что этические проблемы обеих сфер касались ценности человеческой жизни, уважения и достоинства пользователя.

Проблемы были описаны, проанализированы, к прикладным проблемам были сформулированы этические рекомендации решения этих проблем в виде составления проектов кодексов, принципов, декларативных документов.

Список литературы

1. Аверкин А.Н., Гаазе-Рапопорт М.Г., Поспелов Д.А. Толковый словарь по искусственному интеллекту URL:
<http://www.raai.org/library/tolk/aivoc.html#L208> (дата обращения: 27.04.18)
2. Аверкин А.Н., Гаазе-Рапопорт М.Г., Поспелов Д.А. Толковый словарь по искусственному интеллекту URL:
<http://www.raai.org/library/tolk/aivoc.html#L489> (дата обращения: 25.04.18)
3. Азимов А. Я, робот Эксмо; 2010. 1296 с.
4. ГОСТ Р ИСО 8373-2014 Роботы и робототехнические устройства. Термины и определения: п. 2.6 URL:
<http://docs.cntd.ru/document/1200118297> (дата обращения: 10.02.18)
5. Ильичёв Л. Ф., Федосеев П. Н., Ковалёв С. М., Панов В. Г. Философский энциклопедический словарь. — М.: Советская энциклопедия. 1983. 840 с.
6. Коваливский А. Боевые роботы. Конкуренция за будущее //Новый оборонный заказ .Стратегии №5(42) 2016. URL:
<http://dfnc.ru/robot/boevye-roboty-konkurenciya-za-budushhee/> (дата обращения: 09.04.18)
7. Минделл Д. Восстание машин отменяется! Мифы о роботизации: «Альпина нон-фикшн»; 2017 . 310 с.
8. Постановление Правительства РФ от 11.03.2010 N 138 (ред. от 12.07.2016) «Об утверждении Федеральных правил использования воздушного пространства Российской Федерации» URL:
<http://www.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=201851#0> (дата обращения: 04.08.18)
9. Сэндел М. Что такое справедливость «Манн, Иванов и Фербер»: Москва; 2013. 337 с.

10. Теплов О. Железные всадники //GEO 2006 URL:
<http://guide.travel.ru/qatar/124106.html> (дата обращения: 08.05.2018)
11. Чапек К. R.U.R. 104 с. URL: <http://lib.ru/SOCFANT/CHAPEK/rur.txt>
(дата обращения: 16.03.18)
12. Allen C., Wallach W., Smit I. Why Machine Ethics?// IEEE Intelligent Systems, vol. 21, no. 4, (July-Aug. 2006) pp. 12-17.
13. Anderson M., Anderson S.L., and C. Armen Towards Machine Ethics: Implementing Two Action-Based Ethical Theories// Machine Ethics. AAAI Press, (2005) pp. 1–7.
14. Demography report 2010. Older, more numerous and diverse Europeans. European Commission, March 2011. URL:
http://epp.eurostat.ec.europa.eu/cache/ITY_OF FPUB/KE-ET-10-001/EN/KE-ET-10-001- EN.PDF (дата обращения: 09.03.2018)
15. Denning T. Matuszek C. Koscher K. Smith JR. Kohno T. A spotlight on security and privacy risks with future household robots: attacks and lessons.//In: Proceedings of the 11th international conference on Ubiquitous computing 11th international conference on Ubiquitous computing. (2009) pp.105–114.
16. Eurostat: Population structure and ageing URL:
http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing (дата обращения: 09.03.2018)
17. Feil-Seifer D. Mataric M. J. Defining Socially Assistive Robotics. Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics. (2005) USA, pp.465-468.
18. Fong T. Nourbakhsh, I. & Dautenhahn, K. A survey of socially interactive robots.//Robotics and Autonomous Systems (2003) Vol.42, pp.143– 166.

19. International Federation of Robotics URL: <https://ifr.org/service-robots/>
(дата обращения: 08.03.2018)
20. Kitchener K.S. Anderson S.K. Foundations of ethical practice, research, and teaching in psychology and counseling (2011) Routledge, London p. 201.
21. Körtner T. Ethical challenges in the use of social service robots for elderly people//Zeitschrift für Gerontologie und Geriatrie No.4 (2016) pp. 303-306.
22. Moor J.H. The Nature, Importance, and Difficulty of Machine Ethics//IEEE Intelligent Systems, Vol. 21, No. 4 (2006) pp. 19-21.
23. Operto F. Ethics in advanced robotics//IEEE Robot Autom Mag 1 Vol.18 (2011) pp.72–78.
24. Principles of robotics//EPSRC URL:
<https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/> (дата обращения: 14.03.18)
25. Report Of Comest On Robotics Ethics//UNESCO; 2017 pp. 64.
URL :<http://unesdoc.unesco.org/images/0025/002539/253952E.pdf> (дата обращения: 10.03.18)
26. Scassellati B. Admoni, H. and Mataric M. Robots for autism research// Annual Review of Biomedical Engineering, No.14 (2012) pp. 279-280.
27. STRANDS project URL:<http://strands.acin.tuwien.ac.at/> (дата обращения: 29.03.18)
28. Turkle S. Relational artifacts with children and elders: The complexities of cybercompanions//Connection Science18 Vol.4 (2006) pp. 347–361.
29. Wallach.W., Allen C. Smit I. Artificial morality: Top-down, bottom-up, and hybrid approaches//Ethics and Information Technology (2006) pp.149-155.